Douglas C. Montgomery • George C. Runger

APPLIED STATISTICS AND PROBABILITY FOR ENGINEERS

Applied Statistics and Probability for Engineers

Seventh Edition

DOUGLAS C. MONTGOMERY
Arizona State University

GEORGE C. RUNGER
Arizona State University

WILEY

To:
Meredith, Neil, Colin, and Cheryl
Rebecca, Elisa, George and Taylor

Vice President and Director	Laurie Rosatone
Editorial Director	Don Fowley
Development Editor	Ryann Dannelly
Editorial Assistant	Adriana Alecci
Production Editor	Ken Santor
Senior Content Manager	Valerie Zaborski
Product Design Manager	Tom Kulesa
Marketing Development Manager	John LaVacca III
Designer	Joanna Vieira
Photo Editor	Christin Swinford
Production Management Service	Laserwords/SPI
Cover Design	Joanna Vieira

Cover Photo: © Gio_tto/Getty Images

This book was set by SPi Global and printed and bound by LLSC Kendallville.

This book is printed on acid free paper.

Copyright © 2018, 2014, 2011, 2008, 2005, John Wiley & Sons, Inc. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc. 222 Rosewood Drive, Danvers, MA 01923, website www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, (201)748-6011, fax (201)748-6008, website http://www.wiley.com/go/permissions.

Evaluation copies are provided to qualified academics and professionals for review purposes only, for use in their courses during the next academic year. These copies are licensed and may not be sold or transferred to a third party. Upon completion of the review period, please return the evaluation copy to Wiley. Return instructions and a free of charge return shipping label are available at www.wiley.com/go/returnlabel. Outside of the United States, please contact your local representative.

Library of Congress Cataloging-in-Publication Data

Names: Montgomery, Douglas C., author. | Runger, George C., author.
Title: Applied statistics and probability for engineers / Douglas C. Montgomery, Arizona State University, George C. Runger, Arizona State University.
Description: Seventh edition. | Hoboken, NJ : Wiley, [2018] | Includes bibliographical references and index. |
Identifiers: LCCN 2017043465 (print) | LCCN 2017045059 (ebook) | ISBN 9781119400363 (enhanced epub) | ISBN 978-1-119-40030-1
Subjects: LCSH: Statistics. | Probabilities.
Classification: LCC QA276.12 (ebook) | LCC QA276.12 .M645 2018 (print) | DDC 519.5—dc23
LC record available at https://lccn.loc.gov/2017043465

ePub ISBN 978-1-119-40036-3

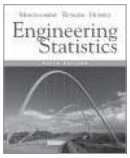
The inside back cover will contain printing identification and country of origin if omitted from this page. In addition, if the ISBN on the back cover differs from the ISBN on this page, the one on the back cover is correct.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Wiley Books by These Authors

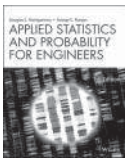
Website: www.wiley.com/college/montgomery



Engineering Statistics, Fifth Edition

by Montgomery and Runger

Introduction to engineering statistics, with topical coverage appropriate for a one-semester course. A modest mathematical level, and an applied approach.



Applied Statistics and Probability for Engineers, Seventh Edition

by Montgomery and Runger

Introduction to engineering statistics, with topical coverage appropriate for eityher a one or two-semester course. An applied approach to solving real-world engineering problems.

Introduction to Statistical Quality Control, Seventh Edition

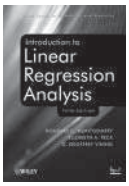
by Douglas C. Montgomery

For a first course in statistical quality control. A comprehensive treatment of statistical methodology for quality control and improvement. Some aspects of quality management are also included such as the six-sigma approach.

Design and Analysis of Experiments, Ninth Edition

by Douglas C. Montgomery

An introduction to design and analysis of expieriments, with the modes prerequisite of a first course in statistical methods. For senior and graduate students for a practitioners, to design and analyze experiments for improving the quality and efficiency of working systems.



Introduction to Linear Regression Analysis,, Fifth Edition

by Mongomery, Peck and Vining

A comprehensive and thoroughly up to date look at regression analysis still most widely used technique in statistics today.

Response Surface Methodology: Process and Product Optimization Using Designed Experiments, Fourth Edition

Website: www.wiley.com/college/myers

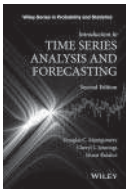
The exploration and optimization of response surfaces, for graduate courses in experimental design and for applied statisticians, engineers and chemical and physical scientists.

Generalized Linear Models: with Applications in Engineering and the Sciences, Second Edition

by Myers, Montgomery and Vining

Website: www.wiley.com/college/myers

An introductory text or reference on generalized linewar models (GLMs). The range of theoretical topics and applications appeals both to students and practicing professionals.



Introduction to Time Series Analysis and Forecasting, Second Edition

by Montgomery, Jennings and Kulahci

Methods for modeling and analyzing times series data, to draw inferences about the datat and generate forecasts useful to he decision maker. Minitab and SAS are used to illustrate how the methods are implemented in practice. For Advanced undergrad/first-year graduate, with a prerequisite of basic statistical methods. Portions of the book require calculus and matrix algebra.

Preface

INTENDED AUDIENCE

This is an introductory textbook for a first course in applied statistics and probability for undergraduate students in engineering and the physical or chemical sciences. These individuals play a significant role in designing and developing new products and manufacturing systems and processes, and they also improve existing systems. Statistical methods are an important tool in these activities because they provide the engineer with both descriptive and analytical methods for dealing with the variability in observed data. Although many of the methods we present are fundamental to statistical analysis in other disciplines, such as business and management, the life sciences, and the social sciences, we have elected to focus on an engineering-oriented audience. We believe that this approach will best serve students in engineering and the chemical/physical sciences and will allow them to concentrate on the many applications of statistics in these disciplines. We have worked hard to ensure that our examples and exercises are engineering- and science-based, and in almost all cases we have used examples of real data—either taken from a published source or based on our consulting experiences.

We believe that engineers in all disciplines should take at least one course in statistics. Unfortunately, because of other requirements, most engineers will only take one statistics course. This book can be used for a single course, although we have provided enough material for two courses in the hope that more students will see the important applications of statistics in their everyday work and elect a second course. We believe that this book will also serve as a useful reference.

We have retained the relatively modest mathematical level of the first five editions. We have found that engineering students who have completed one or two semesters of calculus and have some knowledge of matrix algebra should have no difficulty reading all of the text. It is our intent to give the reader an understanding of the methodology and how to apply it, not the mathematical theory. We have made many enhancements in this edition, including reorganizing and rewriting major portions of the book and adding a number of new exercises.

ORGANIZATION OF THE BOOK

Perhaps the most common criticism of engineering statistics texts is that they are too long. Both instructors and students complain that it is impossible to cover all of the topics in the book in one or even two terms. For authors, this is a serious issue because there is great variety in both the content and level of these courses, and the decisions about what material to delete without limiting the value of the text are not easy. Decisions about which topics to include in this edition were made based on a survey of instructors.

Chapter 1 is an introduction to the field of statistics and how engineers use statistical methodology as part of the engineering problem-solving process. This chapter also introduces the reader to some engineering applications of statistics, including building empirical models, designing engineering experiments, and monitoring manufacturing processes. These topics are discussed in more depth in subsequent chapters.

Chapters 2, 3, 4, and 5 cover the basic concepts of probability, discrete and continuous random variables, probability distributions, expected values, joint probability distributions, and independence. We have given a reasonably complete treatment of these topics but have avoided many of the mathematical or more theoretical details.

Chapter 6 begins the treatment of statistical methods with random sampling; data summary and description techniques, including stem-and-leaf plots, histograms, box plots, and probability plotting; and several types of time series plots. Chapter 7 discusses sampling distributions, the

central limit theorem, and point estimation of parameters. This chapter also introduces some of the important properties of estimators, the method of maximum likelihood, the method of moments, and Bayesian estimation.

Chapter 8 discusses interval estimation for a single sample. Topics included are confidence intervals for means, variances or standard deviations, proportions, prediction intervals, and tolerance intervals. Chapter 9 discusses hypothesis tests for a single sample. Chapter 10 presents tests and confidence intervals for two samples. This material has been extensively rewritten and reorganized. There is detailed information and examples of methods for determining appropriate sample sizes. We want the student to become familiar with how these techniques are used to solve real-world engineering problems and to get some understanding of the concepts behind them. We give a logical, heuristic development of the procedures rather than a formal, mathematical one. We have also included some material on nonparametric methods in these chapters.

Chapters 11 and 12 present simple and multiple linear regression including model adequacy checking and regression model diagnostics and an introduction to logistic regression. We use matrix algebra throughout the multiple regression material (Chapter 12) because it is the only easy way to understand the concepts presented. Scalar arithmetic presentations of multiple regression are awkward at best, and we have found that undergraduate engineers are exposed to enough matrix algebra to understand the presentation of this material.

Chapters 13 and 14 deal with single- and multifactor experiments, respectively. The notions of randomization, blocking, factorial designs, interactions, graphical data analysis, and fractional factorials are emphasized. Chapter 15 introduces statistical quality control, emphasizing the control chart and the fundamentals of statistical process control.

WHAT'S NEW IN THIS EDITION

It is highly gratifying that the sixth edition of the text was the most widely used edition in its history. For this seventh edition, we focused on revisions and enhancements aimed at *lowering costs to the students, better engaging them in the learning process, and providing increased support for instructors.*

Lower-Cost Product Formats

According to the publishing industry's most trusted market research firm, Outsell, 2015 represented a turning point in text usage: for the first time, student preference for digital learning materials was higher than for print, and the increase in preference for digital has been accelerating since.

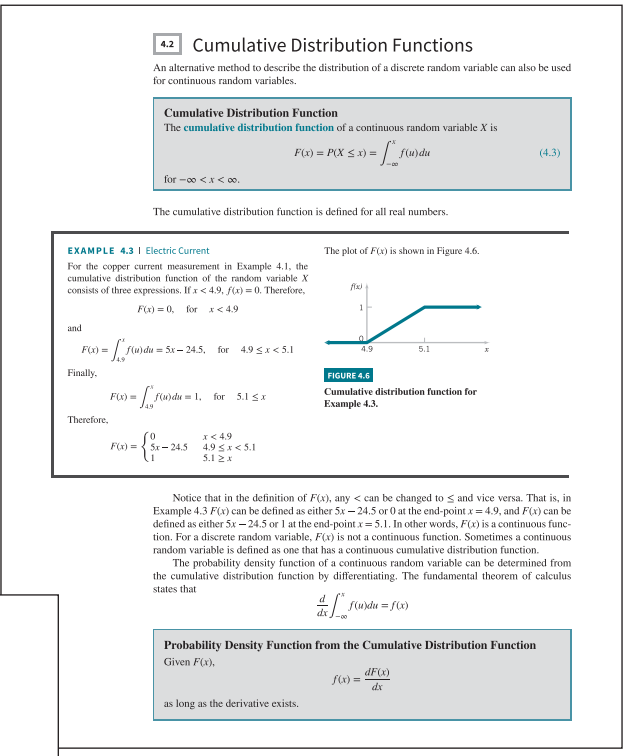
While print remains important for many students as a pedagogical tool, the seventh edition is being delivered in forms that emphasize support for learning from digital materials. *All forms we are providing dramatically reduce the cost to students compared to the sixth edition.* These forms are:

- **WileyPLUS stand-alone.** The WileyPLUS course incorporates the full text and delivers a wealth of integrated, interactive learning aids. See the detailed list of the learning aids below.
- **Stand-alone e-text now with significant enhancements.** The e-text format for the seventh edition adds approximately 100 practice problems with solutions, several hundred basic "Check Your Understanding" exercises for student self-quizzing, and all of the videos included in the WileyPLUS course.
- The two electronic formats listed above can be bundled with an **Abridged Print Companion** that includes a loose-leaf copy of the entire text minus the exercises for a nominal additional cost over the any of the value-priced forms listed above.

To explore these options further or to discuss other options, contact your Wiley account manager (www.wiley.com/go/whosmyrep) or visit the product information page for this text on wiley.com. Instructors may also order bound print companions for the bundled package by contacting their Wiley account representative.

FEATURED IN THIS BOOK

Definitions, Key Concepts, and Equations
Throughout the text, definitions and key concepts and equations are highlighted by a box to emphasize their importance.



4.2 Cumulative Distribution Functions

An alternative method to describe the distribution of a discrete random variable can also be used for continuous random variables.

Cumulative Distribution Function
The cumulative distribution function of a continuous random variable X is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)\,du \qquad (4.3)$$

for $-\infty < x < \infty$.

The cumulative distribution function is defined for all real numbers.

EXAMPLE 4.3 | Electric Current

For the copper current measurement in Example 4.1, the cumulative distribution function of the random variable X consists of three expressions. If $x < 4.9$, $f(x) = 0$. Therefore,

$$F(x) = 0, \quad \text{for} \quad x < 4.9$$

and

$$F(x) = \int_{4.9}^{x} f(u)\,du = 5x - 24.5, \quad \text{for} \quad 4.9 \le x < 5.1$$

Finally,

$$F(x) = \int_{4.9}^{x} f(u)\,du = 1, \quad \text{for} \quad 5.1 \le x$$

Therefore,

$$F(x) = \begin{cases} 0 & x < 4.9 \\ 5x - 24.5 & 4.9 \le x < 5.1 \\ 1 & 5.1 \ge x \end{cases}$$

The plot of $F(x)$ is shown in Figure 4.6.

FIGURE 4.6
Cumulative distribution function for Example 4.3.

Notice that in the definition of $F(x)$, any $<$ can be changed to $\le$ and vice versa. That is, in Example 4.3 $F(x)$ can be defined as either $5x - 24.5$ or 0 at the end-point $x = 4.9$, and $F(x)$ can be defined as either $5x - 24.5$ or 1 at the end-point $x = 5.1$. In other words, $F(x)$ is a continuous function. For a discrete random variable, $F(x)$ is not a continuous function. Sometimes a continuous random variable is defined as one that has a continuous cumulative distribution function.

The probability density function of a continuous random variable can be determined from the cumulative distribution function by differentiating. The fundamental theorem of calculus states that

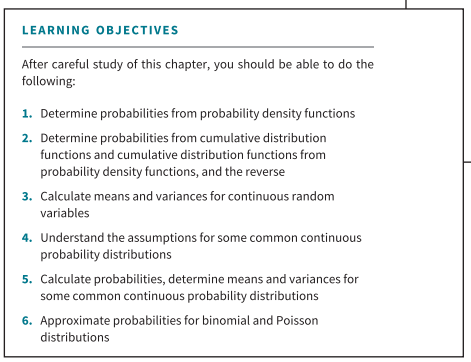
$$\frac{d}{dx}\int_{-\infty}^{x} f(u)du = f(x)$$

Probability Density Function from the Cumulative Distribution Function
Given $F(x)$,

$$f(x) = \frac{dF(x)}{dx}$$

as long as the derivative exists.

Learning Objectives
Learning Objectives at the start of each chapter guide the students in what they are expected to take away from this chapter and serve as a study reference.

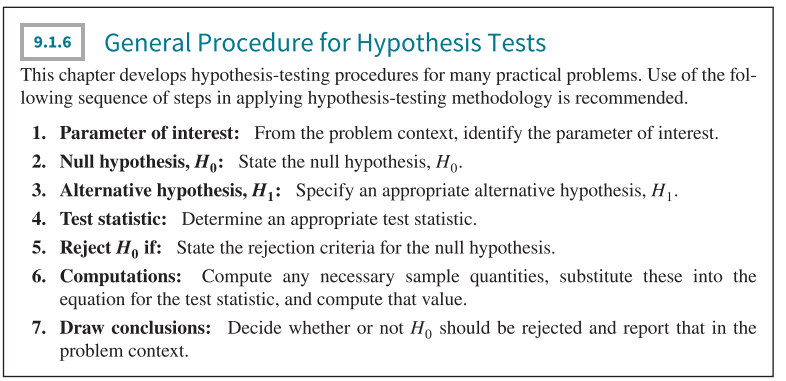


LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Determine probabilities from probability density functions
2. Determine probabilities from cumulative distribution functions and cumulative distribution functions from probability density functions, and the reverse
3. Calculate means and variances for continuous random variables
4. Understand the assumptions for some common continuous probability distributions
5. Calculate probabilities, determine means and variances for some common continuous probability distributions
6. Approximate probabilities for binomial and Poisson distributions

Seven-Step Procedure for Hypothesis Testing
The text introduces a sequence of seven steps in applying hypothesis-testing methodology and explicitly exhibits this procedure in examples.



9.1.6 General Procedure for Hypothesis Tests

This chapter develops hypothesis-testing procedures for many practical problems. Use of the following sequence of steps in applying hypothesis-testing methodology is recommended.

1. **Parameter of interest:** From the problem context, identify the parameter of interest.
2. **Null hypothesis, H_0:** State the null hypothesis, H_0.
3. **Alternative hypothesis, H_1:** Specify an appropriate alternative hypothesis, H_1.
4. **Test statistic:** Determine an appropriate test statistic.
5. **Reject H_0 if:** State the rejection criteria for the null hypothesis.
6. **Computations:** Compute any necessary sample quantities, substitute these into the equation for the test statistic, and compute that value.
7. **Draw conclusions:** Decide whether or not H_0 should be rejected and report that in the problem context.

Figures

Numerous figures throughout the text illustrate statistical concepts in multiple formats.

TABLE 11.1 Oxygen and Hydrocarbon Levels

Observation Number	Hydrocarbon Level x(%)	Purity y(%)
1	0.99	90.01
2	1.02	89.05
3	1.15	91.43
4	1.29	93.74
5	1.46	96.73
6	1.36	94.45
7	0.87	87.59
8	1.23	91.77
9	1.55	99.42
10	1.40	93.65
11	1.19	93.54
12	1.15	92.52

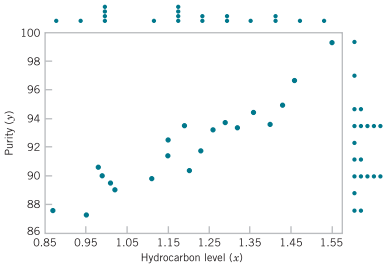


FIGURE 11.1
Scatter diagram of oxygen purity versus hydrocarbon level from Table 11.1.

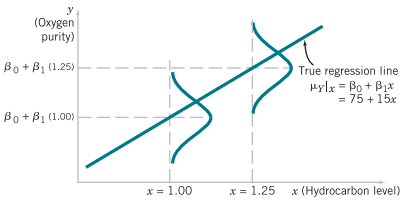


FIGURE 11.2
The distribution of Y for a given value of x for the oxygen purity-hydrocarbon data.

Computer Output

Example throughout the book, use computer output to illustrate the role of modern statistical software.

Stem-and-leaf of Strength

N = 80	Stem	Unit = 1.0
1	7	6
2	8	7
3	9	7
5	10	1 5
8	11	0 5 8
11	12	0 1 3
17	13	1 3 3 4 5 5
25	14	1 2 3 5 6 8 9 9
37	15	0 0 1 3 4 4 6 7 8 8 8 8
(10)	16	0 0 0 3 3 5 7 7 8 9
33	17	0 1 1 2 4 4 5 6 6 8
23	18	0 0 1 1 3 4 6
16	19	0 3 4 6 9 9
10	20	0 1 7 8
6	21	8
5	22	1 8 9
2	23	7
1	24	5

FIGURE 6.5
A typical computer-generated stem-and-leaf diagram.

Example Problems

A set of example problems provides the student with detailed solutions and comments for interesting, real-world situations. Brief practical interpretations have been added in this edition.

EXAMPLE 10.1 | Paint Drying Time

A product developer is interested in reducing the drying time of a primer paint. Two formulations of the paint are tested; formulation 1 is the standard chemistry, and formulation 2 has a new drying ingredient that should reduce the drying time. From experience, it is known that the standard deviation of drying time is 8 minutes, and this inherent variability should be unaffected by the addition of the new ingredient. Ten specimens are painted with formulation 1, and another 10 specimens are painted with formulation 2; the 20 specimens are painted in random order. The two sample average drying times are $\bar{x}_1 = 121$ minutes and $\bar{x}_2 = 112$ minutes, respectively. What conclusions can the product developer draw about the effectiveness of the new ingredient, using $\alpha = 0.05$?

We apply the seven-step procedure to this problem as follows:

1. **Parameter of interest:** The quantity of interest is the difference in mean drying times, $\mu_1 - \mu_2$, and $\Delta_0 = 0$.
2. **Null hypothesis:** H_0: $\mu_1 - \mu_2 = 0$, or H_0: $\mu_1 = \mu_2$.
3. **Alternative hypothesis:** H_1: $\mu_1 > \mu_2$. We want to reject H_0 if the new ingredient reduces mean drying time.
4. **Test statistic:** The test statistic is

$$z_0 = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

where $\sigma_1^2 = \sigma_2^2 = (8)^2 = 64$ and $n_1 = n_2 = 10$.

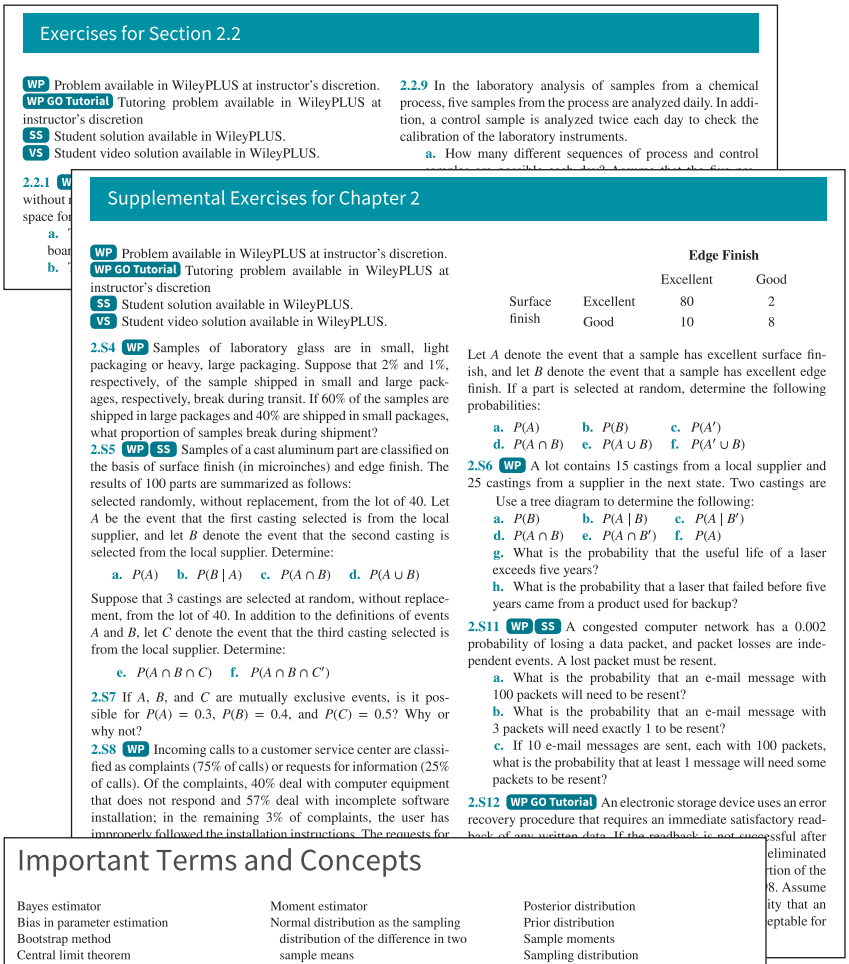
5. **Reject H_0 if:** Reject H_0: $\mu_1 = \mu_2$ if the P-value is less than 0.05.
6. **Computations:** Because $\bar{x}_1 = 121$ minutes and $\bar{x}_2 = 112$ minutes, the test statistic is

$$z_0 = \frac{121 - 112}{\sqrt{\frac{(8)^2}{10} + \frac{(8)^2}{10}}} = 2.52$$

7. **Conclusion:** Because $z_0 = 2.52$, the P-value is $P = 1 - \Phi(2.52) = 0.0059$, so we reject H_0 at the $\alpha = 0.05$ level.

Practical Interpretation: We conclude that adding the new ingredient to the paint significantly reduces the drying time. This is a strong conclusion.

Exercises
Each chapter has an extensive collection of exercises, including **end-of-section exercises** that emphasize the material in that section, **supplemental exercises** at the end of the chapter that cover the scope of chapter topics and require the student to make a decision about the approach they will use to solve the problem, and **mind-expanding exercises** that often require the student to extend the text material somewhat or to apply it in a novel situation. Answers are provided to most odd-numbered exercises in Appendix C in the text, and the *WileyPLUS* online learning environment includes for students complete detailed solutions to selected exercises.



Exercises for Section 2.2

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

2.2.1 W
without r
space for
a.
boar
b.

2.2.9 In the laboratory analysis of samples from a chemical process, five samples from the process are analyzed daily. In addition, a control sample is analyzed twice each day to check the calibration of the laboratory instruments.
a. How many different sequences of process and control

Supplemental Exercises for Chapter 2

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

2.S4 WP Samples of laboratory glass are in small, light packaging or heavy, large packaging. Suppose that 2% and 1%, respectively, of the sample shipped in small and large packages, respectively, break during transit. If 60% of the samples are shipped in large packages and 40% are shipped in small packages, what proportion of samples break during shipment?

2.S5 WP SS Samples of a cast aluminum part are classified on the basis of surface finish (in microinches) and edge finish. The results of 100 parts are summarized as follows:

		Edge Finish	
		Excellent	Good
Surface finish	Excellent	80	2
	Good	10	8

selected randomly, without replacement, from the lot of 40. Let A be the event that the first casting selected is from the local supplier, and let B denote the event that the second casting is selected from the local supplier. Determine:

a. $P(A)$ b. $P(B \mid A)$ c. $P(A \cap B)$ d. $P(A \cup B)$

Suppose that 3 castings are selected at random, without replacement, from the lot of 40. In addition to the definitions of events A and B, let C denote the event that the third casting selected is from the local supplier. Determine:

e. $P(A \cap B \cap C)$ f. $P(A \cap B \cap C')$

Let A denote the event that a sample has excellent surface finish, and let B denote the event that a sample has excellent edge finish. If a part is selected at random, determine the following probabilities:

a. $P(A)$ b. $P(B)$ c. $P(A')$
d. $P(A \cap B)$ e. $P(A \cup B)$ f. $P(A' \cup B)$

2.S6 WP A lot contains 15 castings from a local supplier and 25 castings from a supplier in the next state. Two castings are
Use a tree diagram to determine the following:
a. $P(B)$ b. $P(A \mid B)$ c. $P(A \mid B')$
d. $P(A \cap B)$ e. $P(A \cap B')$ f. $P(A)$
g. What is the probability that the useful life of a laser exceeds five years?
h. What is the probability that a laser that failed before five years came from a product used for backup?

2.S7 If A, B, and C are mutually exclusive events, is it possible for $P(A) = 0.3$, $P(B) = 0.4$, and $P(C) = 0.5$? Why or why not?

2.S8 WP Incoming calls to a customer service center are classified as complaints (75% of calls) or requests for information (25% of calls). Of the complaints, 40% deal with computer equipment that does not respond and 57% deal with incomplete software installation; in the remaining 3% of complaints, the user has improperly followed the installation instructions. The requests for

2.S11 WP SS A congested computer network has a 0.002 probability of losing a data packet, and packet losses are independent events. A lost packet must be resent.
a. What is the probability that an e-mail message with 100 packets will need to be resent?
b. What is the probability that an e-mail message with 3 packets will need exactly 1 to be resent?
c. If 10 e-mail messages are sent, each with 100 packets, what is the probability that at least 1 message will need some packets to be resent?

2.S12 WP GO Tutorial An electronic storage device uses an error recovery procedure that requires an immediate satisfactory read-

eliminated
tion of the
8. Assume
ity that an
eptable for

Important Terms and Concepts

Bayes estimator
Bias in parameter estimation
Bootstrap method
Central limit theorem
Estimator versus estimate
Likelihood function
Maximum likelihood estimator
Mean squared error of an estimator
Minimum variance unbiased estimator
Moment estimator
Normal distribution as the sampling distribution of the difference in two sample means
Normal distribution as the sampling distribution of a sample mean
Parameter estimation
Point estimator
Population or distribution moments
Posterior distribution
Prior distribution
Sample moments
Sampling distribution
Standard error and estimated standard error of an estimator
Statistic
Statistical inference
Unbiased estimator

Important Terms and Concepts
At the end of each chapter is a list of important terms and concepts for an easy self-check and study tool.

Meticulous Ensuring of Accuracy

There is always room for improvement in this area for richly mathematical texts. We appreciate the input of sixth edition adopters and made achieving the greatest possible accuracy of every element of the seventh edition a very high-priority objective. We have been rigorous in reviewing the retained and new content in the text and reviewing and testing the gradable exercises in the WileyPLUS course. We thank the colleagues whom Wiley engaged to execute these activities: (1) the entire text, including all exposition, examples, exercises, and answers have been read and checked for accuracy, including calculations; (2) answers and solutions to all exercises have been checked via independent solution of problems; (3) all questions and exercises in the WileyPLUS course have been checked with a focus on verifying the answers, correcting any conflicting input instructions, and improving the precision of the numerical answers.

Increased Support for Student Engagement

The rapidly increasing use of approaches like flipped classrooms and online course delivery has motivated us to enhance the features for student self-study, especially those that can better engage

students. The following new and expanded interactive features are available in both the WileyPLUS course and the interactive e-text:

- *Video-based examples and problem walk-throughs.* There are now over 100 video solutions and about 40 example videos.
- *Check Your Understanding* exercises are placed at the end of each appropriate section in the WileyPLUS e-text. These simple practice exercises allow students to test whether they comprehend the material they have just read. They can also be used, if the instructor prefers, to assess student readiness to apply concepts and techniques. The program contains over 600 Check Your Understanding questions.

Greater Selection of Assignable Exercises

The seventh edition contains over 1800 exercises, an increase of about 25%. Over 1100 of these exercises are available to the students and appear in the instructor desk copy, e-text version of the text, and about 750 of these exercises have been reserved for instructors in WileyPlus. The e-text and WileyPLUS versions have also added about 300 new exercises to the Check Your Understanding program that consists of about 600 questions. These questions are integrated into the appropriate point in the text exposition to provide students with the ability to test their comprehension of the material just presented before proceeding. In the WileyPLUS course the Check Your Understanding exercises are assignable, enabling instructors to assess student readiness.

Instructor-Only and Algorithmic Exercises for More Exercise Security

The wide and inexpensive availability of solutions to all published textbook problems has added a huge challenge for instructors looking to assess learning. Wiley is embarking on a comprehensive strategy to provide original and reconfigured problems that can decrease the proliferation of problem solutions on the Web. As a first step in this direction, the instructor evaluation/desk copy of this text includes about 750 problems that are *reserved for the instructor only—they do not appear in the student version*. These problems are provided in the instructor desk copy for easy reference. All of these reserved questions appear as assignable problems in WileyPLUS. About 40% of these problems are entirely brand new. The other 60% of these problems have been revised from earlier editions and are now only available as assignable problems. The WileyPLUS course also features *algorithmic homework problems*, machine-gradable problems in which each student has unique values for key variables.

Exercises Collected in the Back of the Book in Instructor Desk Copy

Many instructors have requested that we curtail the practice of frequent revisions of the text. At the same time, the demand for fresh and new problems among instructors is quite high. The exercises in the seventh edition are collected in the back of the instructor desk copy as well as the student e-text and WileyPLUS e-text. We have done this in effort to be able to generate entirely new problem collections that can be adopted without altering the contents of the book. We expect that revisions going forward will be confined largely to expansion of learning aids in the digital forms of the text and updates of the problems and to elongate the life of the print text.

STUDENT RESOURCES

- Data Sets Data sets for all examples and exercises in the text. Visit the student section of the book Web site at www.wiley.com/college/montgomery to access these materials.

INSTRUCTOR RESOURCES

The following resources are available only to instructors who adopt the text:

- **Solutions Manual** All solutions to the exercises in the text.
- **Data Sets** Data sets for all examples and exercises in the text.
- **Image Gallery** of **Text Figures**
- **Section on Logistic Regression**

These instructor-only resources are password-protected. Visit the instructor section of the book Web site at www.wiley.com/college/montgomery to register for a password to access these materials.

COMPUTER SOFTWARE

We have used several different packages, including Excel, to demonstrate computer usage. Minitab can be used for most exercises. A student version of Minitab is available as an option to purchase in a set with this text. Student versions of software often do not have all the functionality that full versions do. Consequently, student versions may not support all the concepts presented in this text. If you would like to adopt for your course the set of this text with the student version of Minitab, please contact your local Wiley representative at www.wiley.com/college/rep.

Alternatively, students may find information about how to purchase the professional version of the software for academic use at www.minitab.com.

WileyPLUS

This online teaching and learning environment integrates the **entire digital textbook** with the most effective instructor and student resources to fit every learning style.

With *WileyPLUS*:

- Students achieve concept mastery in a rich, structured environment that's available 24/7.
- Instructors personalize and manage their course more effectively with assessment, assignments, grade tracking, and more.

WileyPLUS can complement your current textbook or replace the printed text altogether.

For Students **Personalize the learning experience**

Different learning styles, different levels of proficiency, different levels of preparation—each of your students is unique. *WileyPLUS* empowers them to take advantage of their individual strengths:

- Students receive timely access to resources that address their demonstrated needs, and get immediate feedback and remediation when needed.
- Integrated, multi-media resources—including audio and visual exhibits, demonstration problems, and much more—provide multiple study-paths to fit each student's learning preferences and encourage more active learning.
- *WileyPLUS* includes many opportunities for self-assessment linked to the relevant portions of the text. Students can take control of their own learning and practice until they master the material.

For Instructors **Personalize the teaching experience**
WileyPLUS empowers you with the tools and resources you need to make your teaching even more effective:

- You can customize your classroom presentation with a wealth of resources and functionality from PowerPoint slides to a database of rich visuals. You can even add your own materials to your *WileyPLUS* course.
- With *WileyPLUS* you can identify those students who are falling behind and intervene accordingly, without having to wait for them to come to office hours.
- *WileyPLUS* simplifies and automates such tasks as student performance assessment, making assignments, scoring student work, keeping grades, and more.

COURSE SYLLABUS SUGGESTIONS

This is a very flexible textbook because instructors' ideas about what should be in a first course on statistics for engineers vary widely, as do the abilities of different groups of students. Therefore, we hesitate to give too much advice, but will explain how we use the book.

We believe that a first course in statistics for engineers should be primarily an applied statistics course, not a probability course. In our one-semester course we cover all of Chapter 1 (in one or two lectures); overview the material on probability, putting most of the emphasis on the normal distribution (six to eight lectures); discuss most of Chapters 6 through 10 on confidence intervals and tests (twelve to fourteen lectures); introduce regression models in Chapter 11 (four lectures); give an introduction to the design of experiments from Chapters 13 and 14 (six lectures); and present the basic concepts of statistical process control, including the Shewhart control chart from Chapter 15 (four lectures). This leaves about three to four periods for exams and review. Let us emphasize that the purpose of this course is to introduce engineers to how statistics can be used to solve real-world engineering problems, not to weed out the less mathematically gifted students. This course is not the "baby math-stat" course that is all too often given to engineers.

If a second semester is available, it is possible to cover the entire book, including much of the supplemental material, if appropriate for the audience. It would also be possible to assign and work many of the homework problems in class to reinforce the understanding of the concepts. Obviously, multiple regression and more design of experiments would be major topics in a second course.

USING THE COMPUTER

In practice, engineers use computers to apply statistical methods to solve problems. Therefore, we strongly recommend that the computer be integrated into the class. Throughout the book we have presented typical example of the output that can be obtained with modern statistical software. In teaching, we have used a variety of software packages, including Minitab, Statgraphics, JMP, and Statistica. We did not clutter up the book with operational details of these different packages because how the instructor integrates the software into the class is ultimately more important than which package is used. All text data are available in electronic form on the textbook Web site. We have marked these problems with a special icon in the margin.

In our own classrooms, we use the computer in almost every lecture and demonstrate how the technique is implemented in software as soon as it is discussed in the lecture. Many educational institutions have file licenses for statistics software and students can access it for class use. Student versions of many statistical software packages are available at low cost, and students can either purchase their own copy or use the products available through the institution. We have found that greatly improves the pace of the course and student understanding of the material.

Users should be aware that final answers may differ slightly due to different numerical precision and rounding protocols among softwares.

ACKNOWLEDGMENTS

We would like to express our grateful appreciation to the many organizations and individuals who have contributed to this book. Many instructors who used the previous editions provided excellent suggestions that we have tried to incorporate in this revision.

We would like to thank the following who assisted in contributing to and/or reviewing material for the *WileyPLUS* course:

Michael DeVasher, *Rose-Hulman Institute of Technology*

Craig Downing, *Rose-Hulman Institute of Technology*

Julie Fortune, *University of Alabama in Huntsville*

Rubin Wei, *Texas A&M University*

We would also like to thank the following for their assistance in checking the accuracy and completeness of the exercises and the solutions to exercises.

Dr. Abdelaziz Berrado

Dr. Connie Borror

Aysegul Demirtas

Kerem Demirtas

Patrick Egbunonu, Sindhura Gangu

James C. Ford

Dr. Alejandro Heredia-Langner

Dr. Jing Hu

Dr. Busaba Laungrungrong

Dr. Fang Li

Dr. Nuttha Lurponglukana

Sarah StreettYolande Tra

Dr. Lora Zimmer

We are also indebted to Dr. Smiley Cheng for permission to adapt many of the statistical tables from his excellent book (with Dr. James Fu), *Statistical Tables for Classroom and Exam Room*. Wiley, Prentice Hall, the Institute of Mathematical Statistics, and the editors of Biometrics allowed us to use copyrighted material, for which we are grateful.

Douglas C. Montgomery

George C. Runger

Contents

1 The Role of Statistics in Engineering 1

1.1 The Engineering Method and Statistical Thinking 2
- 1.1.1 Variability 3
- 1.1.2 Populations and Samples 5

1.2 Collecting Engineering Data 5
- 1.2.1 Basic Principles 5
- 1.2.2 Retrospective Study 5
- 1.2.3 Observational Study 6
- 1.2.4 Designed Experiments 6
- 1.2.5 Observing Processes Over Time 9

1.3 Mechanistic and Empirical Models 12

1.4 Probability and Probability Models 15

2 Probability 17

2.1 Sample Spaces and Events 18
- 2.1.1 Random Experiments 18
- 2.1.2 Sample Spaces 19
- 2.1.3 Events 21

2.2 Counting Techniques 23

2.3 Interpretations and Axioms of Probability 26

2.4 Unions of Events and Addition Rules 29

2.5 Conditional Probability 31

2.6 Intersections of Events and Multiplication and Total Probability Rules 34

2.7 Independence 36

2.8 Bayes' Theorem 39

2.9 Random Variables 40

3 Discrete Random Variables and Probability Distributions 42

3.1 Probability Distributions and Probability Mass Functions 43

3.2 Cumulative Distribution Functions 45

3.3 Mean and Variance of a Discrete Random Variable 47

3.4 Discrete Uniform Distribution 49

3.5 Binomial Distribution 51

3.6 Geometric and Negative Binomial Distributions 55

3.7 Hypergeometric Distribution 59

3.8 Poisson Distribution 63

4 Continuous Random Variables and Probability Distributions 66

4.1 Probability Distributions and Probability Density Functions 67

4.2 Cumulative Distribution Functions 70

4.3 Mean and Variance of a Continuous Random Variable 71

4.4 Continuous Uniform Distribution 72

4.5 Normal Distribution 73

4.6 Normal Approximation to the Binomial and Poisson Distributions 79

4.7 Exponential Distribution 83

4.8 Erlang and Gamma Distributions 86

4.9 Weibull Distribution 89

4.10 Lognormal Distribution 90

4.11 Beta Distribution 92

5 Joint Probability Distributions 95

5.1 Joint Probability Distributions for Two Random Variables 96

5.2 Conditional Probability Distributions and Independence 102

5.3 Joint Probability Distributions for More Than Two Random Variables 107

5.4 Covariance and Correlation 110

5.5 Common Joint Distributions 113
- 5.5.1 Multinomial Probability Distribution 113
- 5.5.2 Bivariate Normal Distribution 115

5.6 Linear Functions of Random Variables 117

5.7 General Functions of Random Variables 120

5.8 Moment-Generating Functions 121

6 Descriptive Statistics 126

6.1 Numerical Summaries of Data 127

6.2 Stem-and-Leaf Diagrams 131

6.3 Frequency Distributions and Histograms 135
6.4 Box Plots 139
6.5 Time Sequence Plots 140
6.6 Scatter Diagrams 142
6.7 Probability Plots 144

7 Point Estimation of Parameters and Sampling Distributions 148

7.1 Point Estimation 149
7.2 Sampling Distributions and the Central Limit Theorem 150
7.3 General Concepts of Point Estimation 156
7.3.1 Unbiased Estimators 156
7.3.2 Variance of a Point Estimator 157
7.3.3 Standard Error: Reporting a Point Estimate 158
7.3.4 Bootstrap Standard Error 159
7.3.5 Mean Squared Error of an Estimator 160
7.4 Methods of Point Estimation 161
7.4.1 Method of Moments 162
7.4.2 Method of Maximum Likelihood 163
7.4.3 Bayesian Estimation of Parameters 167

8 Statistical Intervals for a Single Sample 170

8.1 Confidence Interval on the Mean of a Normal Distribution, Variance Known 172
8.1.1 Development of the Confidence Interval and Its Basic Properties 172
8.1.2 Choice of Sample Size 175
8.1.3 One-Sided Confidence Bounds 176
8.1.4 General Method to Derive a Confidence Interval 176
8.1.5 Large-Sample Confidence Interval for μ 177
8.2 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown 179
8.2.1 *t* Distribution 180
8.2.2 *t* Confidence Interval on μ 181
8.3 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution 182
8.4 Large-Sample Confidence Interval for a Population Proportion 185
8.5 Guidelines for Constructing Confidence Intervals 188
8.6 Bootstrap Confidence Interval 189
8.7 Tolerance and Prediction Intervals 189
8.7.1 Prediction Interval for a Future Observation 189
8.7.2 Tolerance Interval for a Normal Distribution 191

9 Tests of Hypotheses for a Single Sample 193

9.1 Hypothesis Testing 194
9.1.1 Statistical Hypotheses 194
9.1.2 Tests of Statistical Hypotheses 196
9.1.3 One-Sided and Two-Sided Hypotheses 202
9.1.4 *P*-Values in Hypothesis Tests 203
9.1.5 Connection between Hypothesis Tests and Confidence Intervals 206
9.1.6 General Procedure for Hypothesis Tests 206
9.2 Tests on the Mean of a Normal Distribution, Variance Known 208
9.2.1 Hypothesis Tests on the Mean 208
9.2.2 Type II Error and Choice of Sample Size 211
9.2.3 Large-Sample Test 215
9.3 Tests on the Mean of a Normal Distribution, Variance Unknown 215
9.3.1 Hypothesis Tests on the Mean 215
9.3.2 Type II Error and Choice of Sample Size 220
9.4 Tests on the Variance and Standard Deviation of a Normal Distribution 222
9.4.1 Hypothesis Tests on the Variance 222
9.4.2 Type II Error and Choice of Sample Size 224
9.5 Tests on a Population Proportion 225
9.5.1 Large-Sample Tests on a Proportion 225
9.5.2 Type II Error and Choice of Sample Size 227
9.6 Summary Table of Inference Procedures for a Single Sample 229
9.7 Testing for Goodness of Fit 229
9.8 Contingency Table Tests 232
9.9 Nonparametric Procedures 234
9.9.1 The Sign Test 235
9.9.2 The Wilcoxon Signed-Rank Test 239
9.9.3 Comparison to the *t*-Test 240
9.10 Equivalence Testing 240
9.11 Combining *P*-Values 242

10 Statistical Inference for Two Samples 244

10.1 **Inference on the Difference in Means of Two Normal Distributions, Variances Known 245**
- 10.1.1 Hypothesis Tests on the Difference in Means, Variances Known 247
- 10.1.2 Type II Error and Choice of Sample Size 249
- 10.1.3 Confidence Interval on the Difference in Means, Variances Known 251

10.2 **Inference on the Difference in Means of Two Normal Distributions, Variances Unknown 253**
- 10.2.1 Hypotheses Tests on the Difference in Means, Variances Unknown 253
- 10.2.2 Type II Error and Choice of Sample Size 259
- 10.2.3 Confidence Interval on the Difference in Means, Variances Unknown 260

10.3 **A Nonparametric Test for the Difference in Two Means 261**
- 10.3.1 Description of the Wilcoxon Rank-Sum Test 262
- 10.3.2 Large-Sample Approximation 263
- 10.3.3 Comparison to the *t*-Test 264

10.4 **Paired *t*-Test 264**

10.5 **Inference on the Variances of Two Normal Distributions 268**
- 10.5.1 *F* Distribution 268
- 10.5.2 Hypothesis Tests on the Equity of Two Variances 270
- 10.5.3 Type II Error and Choice of Sample Size 272
- 10.5.4 Confidence Interval on the Ratio of Two Variances 273

10.6 **Inference on Two Population Proportions 273**
- 10.6.1 Large-Sample Tests on the Difference in Population Proportions 274
- 10.6.2 Type II Error and Choice of Sample Size 276
- 10.6.3 Confidence Interval on the Difference in Population Proportions 277

10.7 **Summary Table and Road Map for Inference Procedures for Two Samples 278**

11 Simple Linear Regression and Correlation 280

11.1 **Empirical Models 281**

11.2 **Simple Linear Regression 284**

11.3 **Properties of the Least Squares Estimators 288**

11.4 **Hypothesis Tests in Simple Linear Regression 288**
- 11.4.1 Use of *t*-Tests 289
- 11.4.2 Analysis of Variance Approach to Test Significance of Regression 291

11.5 **Confidence Intervals 292**
- 11.5.1 Confidence Intervals on the Slope and Intercept 292
- 11.5.2 Confidence Interval on the Mean Response 293

11.6 **Prediction of New Observations 295**

11.7 **Adequacy of the Regression Model 296**
- 11.7.1 Residual Analysis 296
- 11.7.2 Coefficient of Determination (R^2) 298

11.8 **Correlation 299**

11.9 **Regression on Transformed Variables 303**

11.10 **Logistic Regression 305**

12 Multiple Linear Regression 310

12.1 **Multiple Linear Regression Model 311**
- 12.1.1 Introduction 311
- 12.1.2 Least Squares Estimation of the Parameters 314
- 12.1.3 Matrix Approach to Multiple Linear Regression 316
- 12.1.4 Properties of the Least Squares Estimators 321

12.2 **Hypothesis Tests in Multiple Linear Regression 322**
- 12.2.1 Test for Significance of Regression 322
- 12.2.2 Tests on Individual Regression Coefficients and Subsets of Coefficients 325

12.3 **Confidence Intervals in Multiple Linear Regression 329**
- 12.3.1 Confidence Intervals on Individual Regression Coefficients 329
- 12.3.2 Confidence Interval on the Mean Response 330

12.4 **Prediction of New Observations 331**

12.5 **Model Adequacy Checking 333**
- 12.5.1 Residual Analysis 333
- 12.5.2 Influential Observations 335

12.6 **Aspects of Multiple Regression Modeling 337**
- 12.6.1 Polynomial Regression Models 337
- 12.6.2 Categorical Regressors and Indicator Variables 339
- 12.6.3 Selection of Variables and Model Building 341
- 12.6.4 Multicollinearity 349

13 Design and Analysis of Single-Factor Experiments: The Analysis of Variance 351

13.1 Designing Engineering Experiments 352
13.2 Completely Randomized Single-Factor Experiment 353
13.2.1 Example: Tensile Strength 353
13.2.2 Analysis of Variance 354
13.2.3 Multiple Comparisons Following the ANOVA 359
13.2.4 Residual Analysis and Model Checking 361
13.2.5 Determining Sample Size 363
13.3 The Random-Effects Model 365
13.3.1 Fixed Versus Random Factors 365
13.3.2 ANOVA and Variance Components 365
13.4 Randomized Complete Block Design 368
13.4.1 Design and Statistical Analysis 368
13.4.2 Multiple Comparisons 372
13.4.3 Residual Analysis and Model Checking 373

14 Design of Experiments with Several Factors 375

14.1 Introduction 376
14.2 Factorial Experiments 378
14.3 Two-Factor Factorial Experiments 382
14.3.1 Statistical Analysis 382
14.3.2 Model Adequacy Checking 386
14.3.3 One Observation per Cell 387
14.4 General Factorial Experiments 388
14.5 2^k Factorial Designs 390
14.5.1 2^2 Design 390
14.5.2 2^k Design for $k \geq 3$ Factors 396
14.6 Single Replicate of the 2^k Design 402
14.7 Addition of Center Points to a 2^k Design 405
14.8 Blocking and Confounding in the 2^k Design 408
14.9 One-Half Fraction of the 2^k Design 413
14.10 Smaller Fractions: The 2^{k-p} Fractional Factorial 418
14.11 Response Surface Methods and Designs 425

15 Statistical Quality Control 434

15.1 Quality Improvement and Statistics 435
15.1.1 Statistical Quality Control 436
15.1.2 Statistical Process Control 436
15.2 Introduction to Control Charts 436
15.2.1 Basic Principles 436
15.2.2 Design of a Control Chart 440
15.2.3 Rational Subgroups 441
15.2.4 Analysis of Patterns on Control Charts 442
15.3 $\bar{X}$ and R or S Control Charts 444
15.4 Control Charts for Individual Measurements 450
15.5 Process Capability 452
15.6 Attribute Control Charts 456
15.6.1 P Chart (Control Chart for Proportions) 456
15.6.2 U Chart (Control Chart for Defects per Unit) 458
15.7 Control Chart Performance 460
15.8 Time-Weighted Charts 462
15.8.1 Exponentially Weighted Moving-Average Control Chart 462
15.8.2 Cumulative Sum Control Chart 465
15.9 Other SPC Problem-Solving Tools 471
15.10 Decision Theory 473
15.10.1 Decision Models 473
15.10.2 Decision Criteria 474
15.11 Implementing SPC 476

APPENDIX A Statistical Tables and Charts A-3
Table I Summary of Common Probability Distributions A-4
Table II Cumulative Binomial Probabilities $P(X \leq x)$ A-5
Table III Cumulative Standard Normal Distribution A-8
Table IV Percentage Points $\chi^2_{\alpha,\nu}$ of the Chi-Squared Distribution A-10
Table V Percentage Points $t_{\alpha,\nu}$ of the t Distribution A-11
Table VI Percentage Points f_{α,ν_1,ν_2} of the F Distribution A-12
Chart VII Operating Characteristic Curves A-17
Table VIII Critical Values for the Sign Test A-26
Table IX Critical Values for the Wilcoxon Signed-Rank Test A-26
Table X Critical Values for the Wilcoxon Rank-Sum Test A-27
Table XI Factors for Constructing Variables Control Charts A-28
Table XII Factors for Tolerance Intervals A-29

APPENDIX B Bibliography A-31

GLOSSARY G-1
EXERCISES P-1
RESERVE PROBLEMS R-1
INDEX I-1

CHAPTER **1**

The Role of Statistics in Engineering

phillyskater/Getty Images

LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Identify the role that statistics can play in the engineering problem-solving process
2. Discuss how variability affects the data collected and used for making engineering decisions
3. Explain the difference between enumerative and analytical studies
4. Discuss the different methods that engineers use to collect data
5. Identify the advantages that designed experiments have in comparison to other methods of collecting engineering data
6. Explain the differences between mechanistic models and empirical models
7. Discuss how probability and probability models are used in engineering and science

CHAPTER OUTLINE

1.1 The Engineering Method and Statistical Thinking

- **1.1.1** Variability
- **1.1.2** Populations and Samples

1.2 Collecting Engineering Data

- **1.2.1** Basic Principles
- **1.2.2** Retrospective Study
- **1.2.3** Observational Study
- **1.2.4** Designed Experiments
- **1.2.5** Observing Processes Over Time

1.3 Mechanistic and Empirical Models

1.4 Probability and Probability Models

Statistics is a science that helps us make decisions and draw conclusions in the presence of variability. For example, civil engineers working in the transportation field are concerned about the capacity of regional highway systems. A typical problem related to transportation would involve data regarding this specific system's number of nonwork, home-based trips, the number of persons per household, and the number of vehicles per household. The objective would be to produce a trip-generation model relating trips to the number of persons per household and the number of vehicles per household. A statistical technique called *regression analysis* can be used to construct this model. The trip-generation model is an important tool for transportation systems planning. Regression methods are among the most widely used statistical techniques in engineering. They are presented in Chapters 11 and 12.

The hospital emergency department (ED) is an important part of the healthcare delivery system. The process by which patients arrive at the ED is highly variable and can depend on the hour of the day and the day of the week, as well as on longer-term cyclical variations. The service process is also highly variable, depending on the types of services that the patients require, the number of patients in the ED, and how the ED is staffed and organized. An ED's capacity is also limited; consequently, some patients experience long waiting times. How long do patients wait, on average? This is an important question for healthcare providers. If waiting times become excessive, some patients will leave without receiving treatment (LWOT). Patients who LWOT do not have their medical concerns addressed and are at risk for further problems and complications. Therefore, another important question is: What proportion of patients LWOT from the ED? These questions can be solved by employing probability models to describe the ED, and from these models very precise estimates of waiting times and the number of patients who LWOT can be obtained. Probability models that can be used to solve these types of problems are discussed in Chapters 2 through 5.

The concepts of probability and statistics are powerful ones and contribute extensively to the solutions of many types of engineering problems. You encounter many examples of these applications in this book.

1.1 The Engineering Method and Statistical Thinking

An engineer is someone who solves problems of interest to society by the efficient application of scientific principles. Engineers accomplish this by either refining an existing product or process or by designing a new product or process that meets customers' needs. The **engineering**, or **scientific**, **method** is the approach to formulating and solving these problems. The steps in the engineering method are as follows:

1. Develop a clear and concise description of the problem.
2. Identify, at least tentatively, the important factors that affect this problem or that may play a role in its solution.
3. Propose a model for the problem, using scientific or engineering knowledge of the phenomenon being studied. State any limitations or assumptions of the model.
4. Conduct appropriate experiments and collect data to test or validate the tentative model or conclusions made in steps 2 and 3.
5. Refine the model on the basis of the observed data.
6. Manipulate the model to assist in developing a solution to the problem.

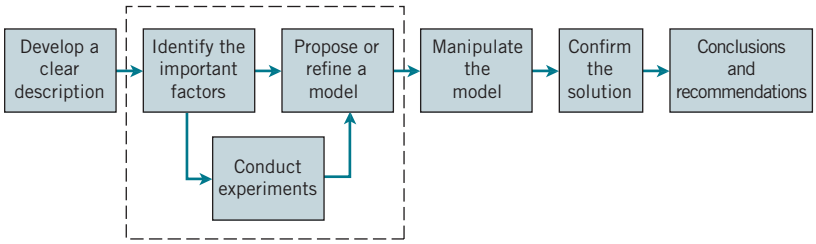


FIGURE 1.1 **The engineering method.**

7. Conduct an appropriate experiment to confirm that the proposed solution to the problem is both effective and efficient.
8. Draw conclusions or make recommendations based on the problem solution.

The steps in the engineering method are shown in Figure 1.1. Many engineering sciences employ the engineering method: the mechanical sciences (statics, dynamics), fluid science, thermal science, electrical science, and the science of materials. Notice that the engineering method features a strong interplay among the problem, the factors that may influence its solution, a model of the phenomenon, and experimentation to verify the adequacy of the model and the proposed solution to the problem. Steps 2–4 in Figure 1.1 are enclosed in a box, indicating that several cycles or iterations of these steps may be required to obtain the final solution. Consequently, engineers must know how to efficiently plan experiments, collect data, analyze and interpret the data, and understand how the observed data relate to the model they have proposed for the problem under study.

The field of **statistics** deals with the collection, presentation, analysis, and use of data to make decisions, solve problems, and design products and processes. In simple terms, **statistics is the science of data**. Because many aspects of engineering practice involve working with data, obviously knowledge of statistics is just as important to an engineer as are the other engineering sciences. Specifically, statistical techniques can be powerful aids in designing new products and systems, improving existing designs, and designing, developing, and improving production processes.

1.1.1 Variability

Statistical methods are used to help us describe and understand **variability**. By variability, we mean that successive observations of a system or phenomenon do not produce exactly the same result. We all encounter variability in our everyday lives, and **statistical thinking** can give us a useful way to incorporate this variability into our decision-making processes. For example, consider the gasoline mileage performance of your car. Do you always get exactly the same mileage performance on every tank of fuel? Of course not—in fact, sometimes the mileage performance varies considerably. This observed variability in gasoline mileage depends on many factors, such as the type of driving that has occurred most recently (city versus highway), the changes in the vehicle's condition over time (which could include factors such as tire inflation, engine compression, or valve wear), the brand and/or octane number of the gasoline used, or possibly even the weather conditions that have been recently experienced. These factors represent potential **sources of variability** in the system. Statistics provides a framework for describing this variability and for learning about which potential sources of variability are the most important or which have the greatest impact on the gasoline mileage performance.

The following example illustrates how we encounter variability in dealing with engineering problems.

EXAMPLE 1.1 | Pull-off Force Variability

Suppose that an engineer is designing a nylon connector to be used in an automotive engine application. The engineer is considering establishing the design specification on wall thickness at 3/32 inch but is somewhat uncertain about the effect of this decision on the connector pull-off force. If the pull-off force is too low, the connector may fail when it is installed in an engine. Eight prototype units are produced and their pull-off forces measured, resulting in the following data (in pounds): 12.6, 12.9, 13.4, 12.3, 13.6, 13.5, 12.6, 13.1. As we anticipated, not all of the prototypes have the same pull-off force. We say that there is variability in the pull-off force measurements.

Because the pull-off force measurements exhibit variability, we consider the pull-off force to be a **random variable**. A convenient way to think of a random variable, say X, that represents a measurement is by using the model

$$X = \mu + \epsilon \tag{1.1}$$

where μ is a constant and ϵ is a random disturbance. The constant remains the same with every measurement, but small changes in the environment, variance in test equipment, differences in the individual parts themselves, and so forth change the value of ϵ. If there were no disturbances, ϵ would always equal zero and X would always be equal to the constant μ. However, this never happens in the real world, so the actual measurements X exhibit variability. We often need to describe, quantify, and ultimately reduce variability.

Figure 1.2 presents a **dot diagram** of these data. The dot diagram is a very useful plot for displaying a small body of data—say, up to about 20 observations. This plot allows us to easily see two features of the data: the **location**, or the middle, and the **scatter** or **variability**. When the number of observations is small, it is usually difficult to identify any specific patterns in the variability, although the dot diagram is a convenient way to see any unusual data features.

The need for statistical thinking arises often in the solution of engineering problems. Consider the engineer designing the connector. From testing the prototypes, he knows that the average pull-off force is 13.0 pounds. However, he thinks that this may be too low for the intended application, so he decides to consider an alternative design with a thicker wall, 1/8 inch in thickness. Eight prototypes of this design are built, and the observed pull-off force measurements are 12.9, 13.7, 12.8, 13.9, 14.2, 13.2, 13.5, and 13.1. The average is 13.4. Results for both samples are plotted as dot diagrams in Figure 1.3. This display gives the impression that increasing the wall thickness has led to an increase in pull-off force. However, there are some obvious questions to ask. For instance, how do we know that another sample of prototypes will not give different results? Is a sample of eight prototypes adequate to give reliable results? If we use the test results obtained so far to conclude that increasing the wall thickness increases the strength, what risks are associated with this decision? For example, is it possible that the apparent increase in pull-off force observed in the thicker prototypes is due only to the inherent variability in the system and that increasing the thickness of the part (and its cost) really has no effect on the pull-off force?

FIGURE 1.2
Dot diagram of the pull-off force data when wall thickness is 3/32 inch.

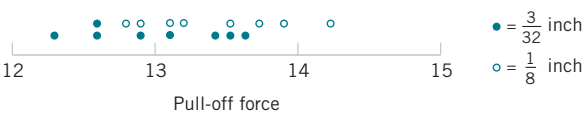


FIGURE 1.3
Dot diagram of pull-off force for two wall thicknesses.

1.1.2 Populations and Samples

Often, physical laws (such as Ohm's law and the ideal gas law) are applied to help design products and processes. We are familiar with this reasoning from general laws to specific cases. But it is also important to reason from a specific set of measurements to more general cases to answer the previous questions. This reasoning comes from a **sample** (such as the eight connectors) to a **population** (such as the connectors that will be in the products that are sold to customers). The reasoning is referred to as **statistical inference**. See Figure 1.4. Historically, measurements were obtained from a sample of people and generalized to a population, and the terminology has remained. Clearly, reasoning based on measurements from some objects to measurements on all objects can result in errors (called *sampling errors*). However, if the sample is selected properly, these risks can be quantified and an appropriate sample size can be determined.

1.2 Collecting Engineering Data

1.2.1 Basic Principles

In the previous subsection, we illustrated some simple methods for summarizing data. Sometimes the data are all of the observations in the population. This results in a **census**. However, in the engineering environment, the data are almost always a **sample** that has been selected from the population. Three basic methods of collecting data are

- A **retrospective study** using historical data
- An **observational study**
- A **designed experiment**

An effective data-collection procedure can greatly simplify the analysis and lead to improved understanding of the population or process that is being studied. We now consider some examples of these data-collection methods.

1.2.2 Retrospective Study

Montgomery, Peck, and Vining (2012) describe an acetone-butyl alcohol distillation column for which concentration of acetone in the distillate (the output product stream) is an important variable. Factors that may affect the distillate are the reboil temperature, the condensate temperature, and the reflux rate. Production personnel obtain and archive the following records:

- The concentration of acetone in an hourly test sample of output product
- The reboil temperature log, which is a record of the reboil temperature over time
- The condenser temperature controller log
- The nominal reflux rate each hour

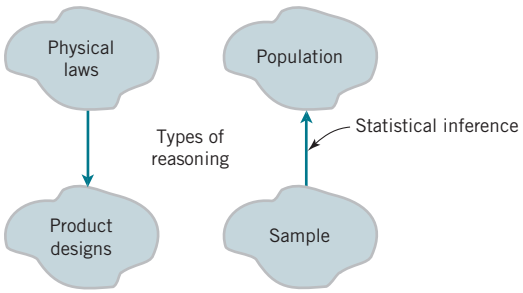


FIGURE 1.4
Statistical inference is one type of reasoning.

The reflux rate should be held constant for this process. Consequently, production personnel change this very infrequently.

A retrospective study would use either all or a sample of the historical process data archived over some period of time. The study objective might be to discover the relationships among the two temperatures and the reflux rate on the acetone concentration in the output product stream. However, this type of study presents some problems:

1. We may not be able to see the relationship between the reflux rate and acetone concentration because the reflux rate did not change much over the historical period.
2. The archived data on the two temperatures (which are recorded almost continuously) do not correspond perfectly to the acetone concentration measurements (which are made hourly). It may not be obvious how to construct an approximate correspondence.
3. Production maintains the two temperatures as closely as possible to desired targets or set points. Because the temperatures change so little, it may be difficult to assess their real impact on acetone concentration.
4. In the narrow ranges within which they do vary, the condensate temperature tends to increase with the reboil temperature. Consequently, the effects of these two process variables on acetone concentration may be difficult to separate.

As you can see, a retrospective study may involve a significant amount of **data**, but those data may contain relatively little useful **information** about the problem. Furthermore, some of the relevant data may be missing, there may be transcription or recording errors resulting in **outliers** (or unusual values), or data on other important factors may not have been collected and archived. In the distillation column, for example, the specific concentrations of butyl alcohol and acetone in the input feed stream are very important factors, but they are not archived because the concentrations are too hard to obtain on a routine basis. As a result of these types of issues, statistical analysis of historical data sometimes identifies interesting phenomena, but solid and reliable explanations of these phenomena are often difficult to obtain.

1.2.3 Observational Study

In an observational study, the engineer observes the process or population, disturbing it as little as possible, and records the quantities of interest. Because these studies are usually conducted for a relatively short time period, sometimes variables that are not routinely measured can be included. In the distillation column, the engineer would design a form to record the two temperatures and the reflux rate when acetone concentration measurements are made. It may even be possible to measure the input feed stream concentrations so that the impact of this factor could be studied.

Generally, an observational study tends to solve problems 1 and 2 and goes a long way toward obtaining accurate and reliable data. However, observational studies may not help resolve problems 3 and 4.

1.2.4 Designed Experiments

In a designed experiment, the engineer makes *deliberate* or *purposeful changes* in the controllable variables of the system or process, observes the resulting system output data, and then makes an inference or decision about which variables are responsible for the observed changes in output performance. The nylon connector example in Section 1.1 illustrates a designed experiment; that is, a deliberate change was made in the connector's wall thickness with the objective of discovering whether or not a stronger pull-off force could be obtained. Experiments designed with basic principles such as **randomization** are needed to establish **cause-and-effect** relationships.

Much of what we know in the engineering and physical-chemical sciences is developed through testing or experimentation. Often engineers work in problem areas in which no scientific

or engineering theory is directly or completely applicable, so experimentation and observation of the resulting data constitute the only way that the problem can be solved. Even when there is a good underlying scientific theory that we may rely on to explain the phenomena of interest, it is almost always necessary to conduct tests or experiments to confirm that the theory is indeed operative in the situation or environment in which it is being applied. Statistical thinking and statistical methods play an important role in planning, conducting, and analyzing the data from engineering experiments. Designed experiments play a very important role in engineering design and development and in the improvement of manufacturing processes.

For example, consider the problem involving the choice of wall thickness for the nylon connector. This is a simple illustration of a designed experiment. The engineer chose two wall thicknesses for the connector and performed a series of tests to obtain pull-off force measurements at each wall thickness. In this simple **comparative experiment**, the engineer is interested in determining whether there is any difference between the 3/32- and 1/8-inch designs. An approach that could be used in analyzing the data from this experiment is to compare the mean pull-off force for the 3/32-inch design to the mean pull-off force for the 1/8-inch design using statistical **hypothesis testing**, which is discussed in detail in Chapters 9 and 10. Generally, a **hypothesis** is a statement about some aspect of the system in which we are interested. For example, the engineer might want to know if the mean pull-off force of a 3/32-inch design exceeds the typical maximum load expected to be encountered in this application, say, 12.75 pounds. Thus, we would be interested in testing the hypothesis that the mean strength exceeds 12.75 pounds. This is called a **single-sample hypothesis-testing problem**. Chapter 9 presents techniques for this type of problem. Alternatively, the engineer might be interested in testing the hypothesis that increasing the wall thickness from 3/32 to 1/8 inch results in an increase in mean pull-off force. It is an example of a **two-sample hypothesis-testing problem**. Two-sample hypothesis-testing problems are discussed in Chapter 10.

Designed experiments offer a very powerful approach to studying complex systems, such as the distillation column. This process has three factors—the two temperatures and the reflux rate—and we want to investigate the effect of these three factors on output acetone concentration. A good experimental design for this problem must ensure that we can separate the effects of all three factors on the acetone concentration. The specified values of the three factors used in the experiment are called **factor levels**. Typically, we use a small number of levels such as two or three for each factor. For the distillation column problem, suppose that we use two levels, "high" and "low" (denoted +1 and −1, respectively), for each of the three factors. A very reasonable experiment design strategy uses every possible combination of the factor levels to form a basic experiment with eight different settings for the process. This type of experiment is called a **factorial experiment**. See Table 1.1 for this experimental design.

TABLE 1.1 **The Designed Experiment (Factorial Design) for the Distillation Column**

Reboil Temp.	Condensate Temp.	Reflux Rate
−1	−1	−1
+1	−1	−1
−1	+1	−1
+1	+1	−1
−1	−1	+1
+1	−1	+1
−1	+1	+1
+1	+1	+1

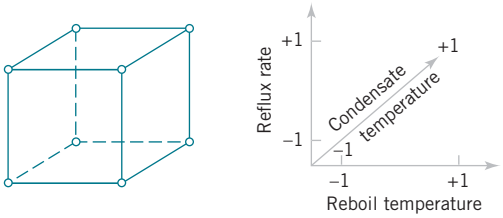


FIGURE 1.5
The factorial design for the distillation column.

Figure 1.5 illustrates that this design forms a cube in terms of these high and low levels. With each setting of the process conditions, we allow the column to reach equilibrium, take a sample of the product stream, and determine the acetone concentration. We then can draw specific inferences about the effect of these factors. Such an approach allows us to proactively study a population or process.

An important advantage of factorial experiments is that they allow one to detect an **interaction** between factors. Consider only the two temperature factors in the distillation experiment. Suppose that the response concentration is poor when the reboil temperature is *low*, regardless of the condensate temperature. That is, the condensate temperature has no effect when the reboil temperature is *low*. However, when the reboil temperature is *high*, a *high* condensate temperature generates a good response, but a *low* condensate temperature generates a poor response. That is, the condensate temperature changes the response when the reboil temperature is *high*. The effect of condensate temperature depends on the setting of the reboil temperature, and these two factors are said to interact in this case. If the four combinations of *high* and *low* reboil and condensate temperatures were not tested, such an interaction would not be detected.

We can easily extend the factorial strategy to more factors. Suppose that the engineer wants to consider a fourth factor, type of distillation column. There are two types: the standard one and a newer design. Figure 1.6 illustrates how all four factors—reboil temperature, condensate temperature, reflux rate, and column design—could be investigated in a factorial design. Because all four factors are still at two levels, the experimental design can still be represented geometrically as a cube (actually, it's a *hypercube*). Notice that as in any factorial design, all possible combinations of the four factors are tested. The experiment requires 16 trials.

Generally, if there are k factors and each has two levels, a factorial experimental design will require 2^k runs. For example, with $k = 4$, the 2^4 design in Figure 1.6 requires 16 tests. Clearly, as the number of factors increases, the number of trials required in a factorial experiment increases rapidly; for instance, eight factors each at two levels would require 256 trials. This quickly becomes unfeasible from the viewpoint of time and other resources. Fortunately, with four to five or more factors, it is usually unnecessary to test all possible combinations of factor levels. A **fractional factorial experiment** is a variation of the basic factorial arrangement in which only a subset of the factor combinations is actually tested. Figure 1.7 shows a fractional factorial experimental design for the distillation column. The circled test combinations in this figure are the only test combinations that need to be run. This experimental design

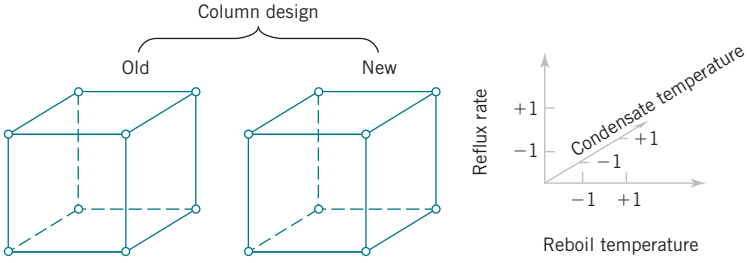


FIGURE 1.6 **A four-factorial experiment for the distillation column.**

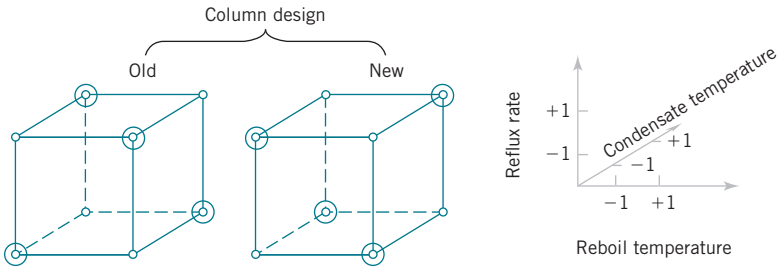


FIGURE 1.7 **A fractional factorial experiment for the distillation column.**

requires only 8 runs instead of the original 16; consequently it would be called a **one-half fraction**. This is an excellent experimental design in which to study all four factors. It will provide good information about the individual effects of the four factors and some information about how these factors interact.

Factorial and fractional factorial experiments are used extensively by engineers and scientists in industrial research and development, where new technology, products, and processes are designed and developed and where existing products and processes are improved. Since so much engineering work involves testing and experimentation, it is essential that all engineers understand the basic principles of planning efficient and effective experiments. We discuss these principles in Chapter 13. Chapter 14 concentrates on the factorial and fractional factorials that we have introduced here.

1.2.5 Observing Processes Over Time

Often data are collected over time. In this case, it is usually very helpful to plot the data versus time in a **time series plot**. Phenomena that might affect the system or process often become more visible in a time-oriented plot and the concept of stability can be better judged.

Figure 1.8 is a dot diagram of acetone concentration readings taken hourly from the distillation column described in Section 1.2.2. The large variation displayed on the dot diagram indicates considerable variability in the concentration, but the chart does not help explain the reason for the variation. The time series plot is shown in Figure 1.9. A shift in the process mean level is visible in the plot and an estimate of the time of the shift can be obtained.

W. Edwards Deming, a very influential industrial statistician, stressed that it is important to understand the nature of variability in processes and systems over time. He conducted an

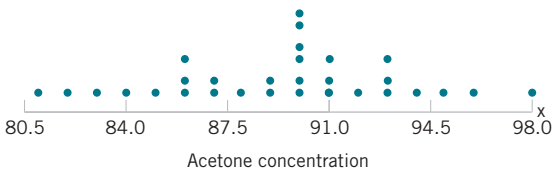


FIGURE 1.8
The dot diagram illustrates variation but does not identify the problem.

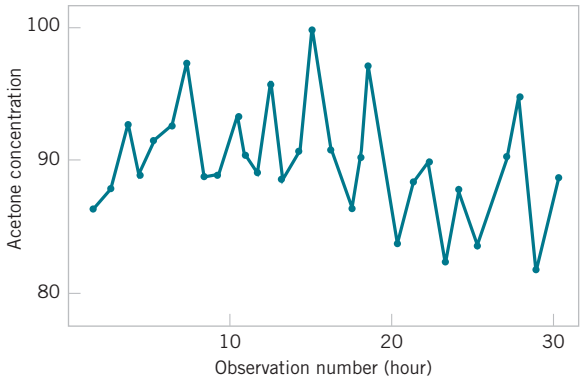


FIGURE 1.9
A time series plot of concentration provides more information than the dot diagram.

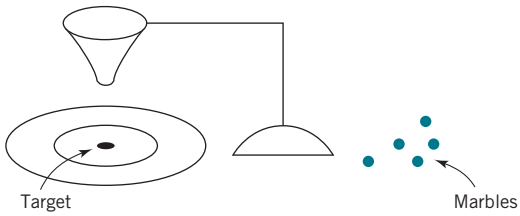


FIGURE 1.10
Deming's funnel experiment.

experiment in which he attempted to drop marbles as close as possible to a target on a table. He used a funnel mounted on a ring stand and the marbles were dropped into the funnel. See Figure 1.10. The funnel was aligned as closely as possible with the center of the target. He then used two different strategies to operate the process. (1) He never moved the funnel. He just dropped one marble after another and recorded the distance from the target. (2) He dropped the first marble and recorded its location relative to the target. He then moved the funnel an equal and opposite distance in an attempt to compensate for the error. He continued to make this type of adjustment after each marble was dropped.

After both strategies were completed, he noticed that the variability of the distance from the target for strategy 2 was approximately twice as large than for strategy 1. The adjustments to the funnel increased the deviations from the target. The explanation is that the error (the deviation of the marble's position from the target) for one marble provides no information about the error that will occur for the next marble. Consequently, adjustments to the funnel do not decrease future errors. Instead, they tend to move the funnel farther from the target.

This interesting experiment points out that adjustments to a process based on random disturbances can actually *increase* the variation of the process. This is referred to as **overcontrol** or **tampering**. Adjustments should be applied only to compensate for a nonrandom shift in the process—then they can help. A computer simulation can be used to demonstrate the lessons of the funnel experiment. Figure 1.11 displays a time plot of 100 measurements (denoted as y) from a process in which only random disturbances are present. The target value for the process is 10 units. The figure displays the data with and without adjustments that are applied to the process mean in an attempt to produce data closer to target. Each adjustment is equal and opposite to the deviation of the previous measurement from target. For example, when the measurement is 11 (one unit above target), the mean is reduced by one unit before the next measurement is generated. The overcontrol increases the deviations from the target.

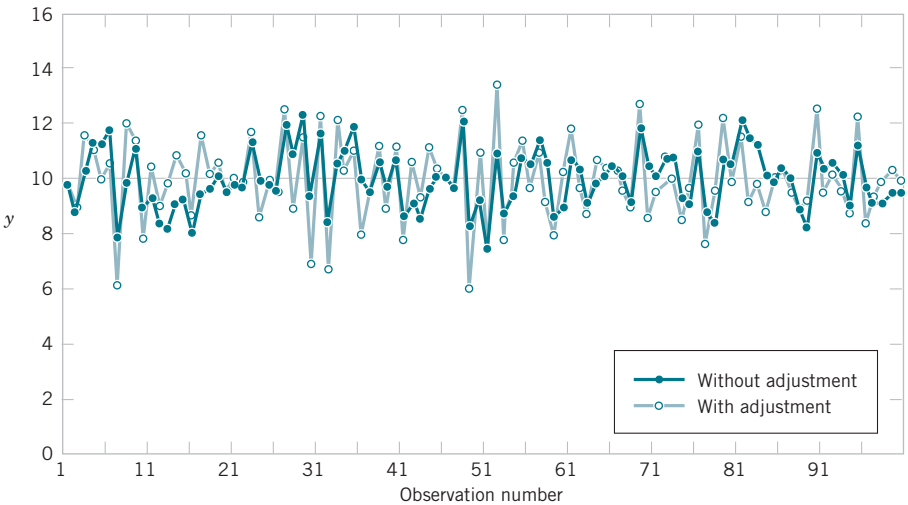


FIGURE 1.11 **Adjustments applied to random disturbances overcontrol the process and increase the deviations from target.**

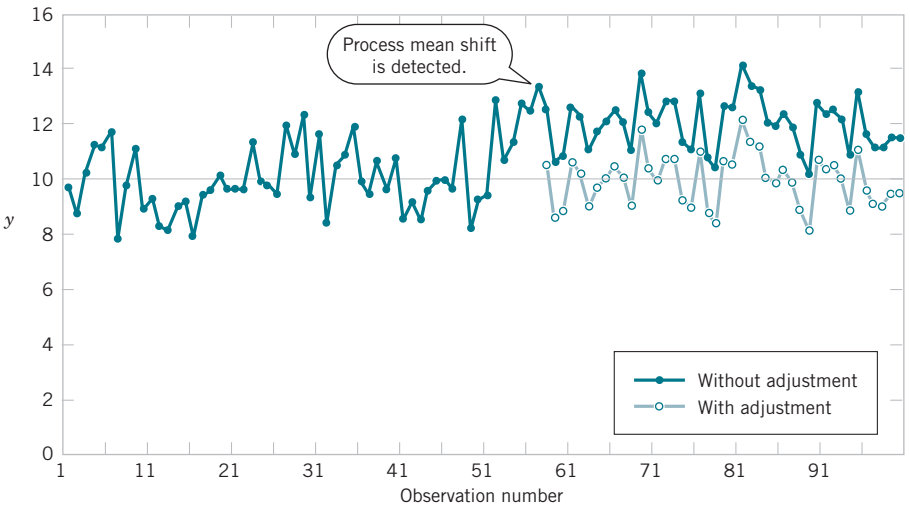


FIGURE 1.12 **Process mean shift is detected at observation number 57, and one adjustment (a decrease of two units) reduces the deviations from target.**

Figure 1.12 displays the data without adjustment from Figure 1.11, except that the measurements after observation number 50 are increased by two units to simulate the effect of a shift in the mean of the process. When there is a true shift in the mean of a process, an adjustment can be useful. Figure 1.12 also displays the data obtained when one adjustment (a decrease of two units) is applied to the mean after the shift is detected (at observation number 57). Note that this adjustment decreases the deviations from target.

The question of when to apply adjustments (and by what amounts) begins with an understanding of the types of variation that affect a process. The use of a **control chart** is an invaluable way to examine the variability in time-oriented data. Figure 1.13 presents a control chart for the concentration data from Figure 1.9. The **center line** on the control chart is just the average of the concentration measurements for the first 20 samples ($\bar{x} = 91.5$ g/l) when the process is stable. The **upper control limit** and the **lower control limit** are a pair of statistically derived limits that reflect the inherent or natural variability in the process. These limits are located 3 standard deviations of the concentration values above and below the center line. If the process is operating as it should without any external sources of variability present in the system, the concentration measurements should fluctuate randomly around the center line, and almost all of them should fall between the control limits.

In the control chart of Figure 1.13, the visual frame of reference provided by the center line and the control limits indicates that some upset or disturbance has affected the process around sample 20 because all of the following observations are below the center line, and two of them actually fall below the lower control limit. This is a very strong signal that corrective action is required in this process. If we can find and eliminate the underlying cause of this upset, we can improve process performance considerably. Thus control limits serve as decision rules about actions that could be taken to improve the process.

Furthermore, Deming pointed out that data from a process are used for different types of conclusions. Sometimes we collect data from a process to evaluate current production. For example, we might sample and measure resistivity on three semiconductor wafers selected from a lot and use this information to evaluate the lot. This is called an **enumerative study**. However, in many cases, we use data from current production to evaluate future production. We apply conclusions to a conceptual, future population. Deming called this an **analytic study**. Clearly this requires an assumption of a **stable** process, and Deming emphasized that control charts were needed to justify this assumption. See Figure 1.14 as an illustration.

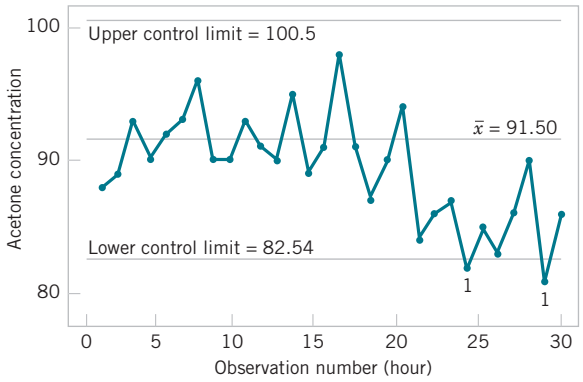


FIGURE 1.13

A control chart for the chemical process concentration data.

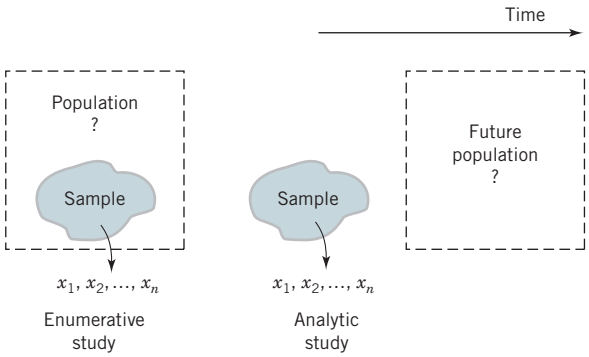


FIGURE 1.14

Enumerative versus analytic study.

The use of control charts is a very important application of statistics for monitoring, controlling, and improving a process. The branch of statistics that makes use of control charts is called **statistical process control**, or **SPC**. We discuss SPC and control charts in Chapter 15.

1.3 Mechanistic and Empirical Models

Models play an important role in the analysis of nearly all engineering problems. Much of the formal education of engineers involves learning about the models relevant to specific fields and the techniques for applying these models in problem formulation and solution. As a simple example, suppose that we are measuring the flow of current in a thin copper wire. Our model for this phenomenon might be Ohm's law:

$$\text{Current} = \text{Voltage}/\text{Resistance}$$

or

$$I = E/R \tag{1.2}$$

We call this type of model a **mechanistic model** because it is built from our underlying knowledge of the basic physical mechanism that relates these variables. However, if we performed this measurement process more than once, perhaps at different times, or even on different days, the observed current could differ slightly because of small changes or variations in factors that are not completely controlled, such as changes in ambient temperature, fluctuations in performance of the gauge, small impurities present at different locations in the wire, and drifts in the voltage source. Consequently, a more realistic model of the observed current might be

$$I = E/R + \epsilon \tag{1.3}$$

where ϵ is a term added to the model to account for the fact that the observed values of current flow do not perfectly conform to the mechanistic model. We can think of ϵ as a term that includes the effects of all unmodeled sources of variability that affect this system.

Sometimes engineers work with problems for which no simple or well-understood mechanistic model explains the phenomenon. For instance, suppose that we are interested in the number average molecular weight (M_n) of a polymer. Now we know that M_n is related to the viscosity of the material (V), and it also depends on the amount of catalyst (C) and the temperature (T) in the polymerization reactor when the material is manufactured. The relationship between M_n and these variables is

$$M_n = f(V, C, T) \tag{1.4}$$

say, where the *form* of the function f is unknown. Perhaps a working model could be developed from a first-order Taylor series expansion, which would produce a model of the form

$$M_n = \beta_0 + \beta_1 V + \beta_2 C + \beta_3 T \tag{1.5}$$

where the β's are unknown parameters. Now just as in Ohm's law, this model will not exactly describe the phenomenon, so we should account for the other sources of variability that may affect the molecular weight by adding another term to the model; therefore,

$$M_n = \beta_0 + \beta_1 V + \beta_2 C + \beta_3 T + \epsilon \tag{1.6}$$

is the model that we will use to relate molecular weight to the other three variables. This type of model is called an **empirical model**; that is, it uses our engineering and scientific knowledge of the phenomenon, but it is not directly developed from our theoretical or first-principles understanding of the underlying mechanism.

To illustrate these ideas with a specific example, consider the data in Table 1.2, which contains data on three variables that were collected in an observational study in a semiconductor manufacturing plant. In this plant, the finished semiconductor is wire-bonded to a frame. The variables reported are pull strength (a measure of the amount of force required to break the bond), the wire length, and the height of the die. We would like to find a model relating pull strength to wire length and die height. Unfortunately, there is no physical mechanism that we can easily apply here, so it does not seem likely that a mechanistic modeling approach will be successful.

Figure 1.15 presents a three-dimensional plot of all 25 observations on pull strength, wire length, and die height. From examination of this plot, we see that pull strength increases as both wire length and die height increase. Furthermore, it seems reasonable to think that a model such as

$$\text{Pull strength} = \beta_0 + \beta_1(\text{wire length}) + \beta_2(\text{die height}) + \epsilon$$

would be appropriate as an empirical model for this relationship. In general, this type of empirical model is called a **regression model**. In Chapters 11 and 12 we show how to build these models and test their adequacy as approximating functions. We use a method for estimating the parameters in regression models, called the method of **least squares**, that traces its origins to work by Karl Gauss. Essentially, this method chooses the parameters in the empirical model (the β's) to minimize the sum of the squared distances in each data point and the plane represented by the model equation. Applying this technique to the data in Table 1.2 results in

$$\widehat{\text{Pull Strength}} = 2.26 + 2.74\ (\text{wire length}) + 0.0125\ (\text{die height}) \tag{1.7}$$

where the "hat," or circumflex, over pull strength indicates that this is an estimated or predicted quality.

Figure 1.16 is a plot of the predicted values of pull strength versus wire length and die height obtained from Equation 1.7. Notice that the predicted values lie on a plane above the wire length–die height space. From the plot of the data in Figure 1.15, this model does not appear

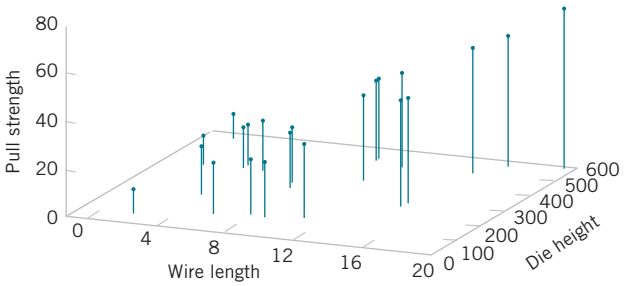


FIGURE 1.15
Three-dimensional plot of the wire bond pull strength data.

TABLE 1.2 Wire Bond Pull Strength Data

Observation Number	Pull Strength y	Wire Length x_1	Die Height x_2
1	9.95	2	50
2	24.45	8	110
3	31.75	11	120
4	35.00	10	550
5	25.02	8	295
6	16.86	4	200
7	14.38	2	375
8	9.60	2	52
9	24.35	9	100
10	27.50	8	300
11	17.08	4	412
12	37.00	11	400
13	41.95	12	500
14	11.66	2	360
15	21.65	4	205
16	17.89	4	400
17	69.00	20	600
18	10.30	1	585
19	34.93	10	540
20	46.59	15	250
21	44.88	15	290
22	54.12	16	510
23	56.63	17	590
24	22.13	6	100
25	21.15	5	400

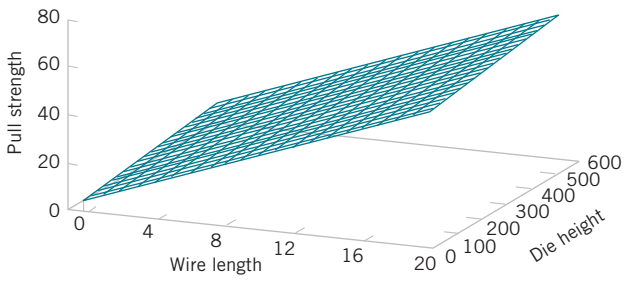


FIGURE 1.16
Plot of predicted values of pull strength from the empirical model.

unreasonable. The empirical model in Equation 1.7 could be used to predict values of pull strength for various combinations of wire length and die height that are of interest. Essentially, an engineer could use the empirical model in exactly the same way as a mechanistic model.

1.4 Probability and Probability Models

Section 1.1 mentioned that decisions often need to be based on measurements from only a subset of objects selected in a sample. This process of reasoning from a sample of objects to conclusions for a population of objects was referred to as *statistical inference*. A sample of three wafers selected from a large production lot of wafers in semiconductor manufacturing was an example mentioned. To make good decisions, an analysis of how well a sample represents a population is clearly necessary. If the lot contains defective wafers, how well will the sample detect these defective items? How can we quantify the criterion to "detect well"? Basically, how can we quantify the risks of decisions based on samples? Furthermore, how should samples be selected to provide good decisions—ones with acceptable risks? **Probability models** help quantify the risks involved in statistical inference, that is, the risks involved in decisions made every day.

More details are useful to describe the role of probability models. Suppose that a production lot contains 25 wafers. If all the wafers are defective or all are good, clearly any sample will generate all defective or all good wafers, respectively. However, suppose that only one wafer in the lot is defective. Then a sample might or might not detect (include) the wafer. A probability model, along with a method to select the sample, can be used to quantify the risks that the defective wafer is or is not detected. Based on this analysis, the size of the sample might be increased (or decreased). The risk here can be interpreted as follows. Suppose that a series of lots, each with exactly one defective wafer, is sampled. The details of the method used to select the sample are postponed until randomness is discussed in the next chapter. Nevertheless, assume that the same-size sample (such as three wafers) is selected in the same manner from each lot. The proportion of the lots in which the defective wafer are included in the sample or, more specifically, the limit of this proportion as the number of lots in the series tends to infinity, is interpreted as the probability that the defective wafer is detected.

A probability model is used to calculate this proportion under reasonable assumptions for the manner in which the sample is selected. This is fortunate because we do not want to attempt to sample from an infinite series of lots. Problems of this type are worked in Chapters 2 and 3. More importantly, this probability provides valuable, quantitative information regarding any decision about lot quality based on the sample.

Recall from Section 1.1 that a population might be conceptual, as in an analytic study that applies statistical inference to future production based on the data from current production. When populations are extended in this manner, the role of statistical inference and the associated probability models become even more important.

In the previous example, each wafer in the sample was classified only as defective or not. Instead, a continuous measurement might be obtained from each wafer. In Section 1.2.5, concentration measurements were taken at periodic intervals from a production process. Figure 1.8 shows that variability is present in the measurements, and there might be concern that the process has moved from the target setting for concentration. Similar to the defective wafer, one might want to quantify our ability to detect a process change based on the sample data. Control limits were mentioned in Section 1.2.5 as decision rules for whether or not to adjust a process. The probability that a particular process change is detected can be calculated with a probability model for concentration measurements. Models for continuous measurements are developed based on plausible assumptions for the data and a result known as the *central limit theorem*, and the associated normal distribution is a particularly valuable probability model for statistical inference. Of course, a check of assumptions is important. These types of probability models are discussed in Chapter 4. The objective is still to quantify the risks inherent in the inference made from the sample data.

Throughout Chapters 6 through 15, we base decisions on statistical inference from sample data. We use continuous probability models, specifically the normal distribution, extensively to quantify the risks in these decisions and to evaluate ways to collect the data and how large a sample should be selected.

Important Terms and Concepts

Analytic study
Cause and effect
Designed experiment
Empirical model
Engineering method
Enumerative study
Factorial experiment
Fractional factorial experiment
Hypothesis
Hypothesis testing
Interaction
Mechanistic model
Observational study
Overcontrol
Population
Probability model
Random variable
Randomization
Retrospective study
Sample
Scientific method
Statistical inference
Statistical process control
Statistical thinking
Tampering
Time series
Variability

CHAPTER 2

Probability

Erik Isakson/Getty Images

LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Understand and describe sample spaces and events for random experiments with graphs, tables, lists, or tree diagrams
2. Interpret probabilities and use the probabilities of outcomes to calculate probabilities of events in discrete sample spaces
3. Use permuations and combinations to count the number of outcomes in both an event and the sample space
4. Calculate the probabilities of joint events such as unions and intersections from the probabilities of individual events
5. Interpret and calculate conditional probabilities of events
6. Determine the independence of events and use independence to calculate probabilities
7. Use Bayes' theorem to calculate conditional probabilities
8. Describe random variables and the difference between continuous and discrete random variables

CHAPTER OUTLINE

2.1 Sample Spaces and Events

- **2.1.1** Random Experiments
- **2.1.2** Sample Spaces
- **2.1.3** Events

2.2 Counting Techniques

2.3 Interpretations and Axioms of Probability

2.4 Unions of Events and Addition Rules

2.5 Conditional Probability

2.6 Intersections of Events and Multiplication and Total Probability Rules

2.7 Independence

2.8 Bayes' Theorem

2.9 Random Variables

An athletic woman in her 20s arrives at the emergency department complaining of dizziness after running in hot weather. An electrocardiogram is used to check for a heart attack, and the patient generates an abnormal result. The test has a false-positive rate 0.1 (the probability of an abnormal result when the patient is normal) and a false-negative rate of 0.1 (the probability of a normal result when the patient is abnormal). Furthermore, it might be assumed that the prior probability of a heart attack for this patient is 0.001. Although the abnormal test is a concern, you might be surprised to learn that the probability of a heart attack given the electrocardiogram result is still less than 0.01. See "Why Clinicians are Natural Bayesians" (2005, *www.bmj.com/content/330/7499/1080*) for details of this example and others.

The key is to properly combine the given probabilities. Furthermore, the exact same analysis used for this medical example can be applied to tests of engineered products. Consequently, knowledge of how to manipulate probabilities in order to assess risks and make better decisions is important throughout scientific and engineering disciplines. In this chapter, the laws of probability are presented and used to assess risks in cases such as this one and numerous others.

2.1 Sample Spaces and Events

2.1.1 Random Experiments

If we measure the current in a thin copper wire, we are conducting an experiment. However, day-to-day repetitions of the measurement can differ slightly because of small variations in variables that are not controlled in our experiment, including changes in ambient temperatures, slight variations in the gauge and small impurities in the chemical composition of the wire (if different locations are selected), and current source drifts. Consequently, this experiment (as well as many we conduct) is said to have a *random* component. In some cases, the random variations are small enough, relative to our experimental goals, that they can be ignored. However, no matter how carefully our experiment is designed and conducted, the variation is almost always present, and its magnitude can be large enough that the important conclusions from our experiment are not obvious. In these cases, the methods presented in this book for modeling and analyzing experimental results are quite valuable.

Our goal is to understand, quantify, and model the type of variations that we often encounter. When we incorporate the variation into our thinking and analyses, we can make informed judgments from our results that are not invalidated by the variation.

Models and analyses that include variation are not different from models used in other areas of engineering and science. Figure 2.1 displays the important components. A mathematical model (or abstraction) of the physical system is developed. It need not be a perfect abstraction. For example, Newton's laws are not perfect descriptions of our physical universe. Still, they are useful models that can be studied and analyzed to approximately quantify the performance of a wide range of engineered products. Given a mathematical abstraction that is validated with measurements from our system, we can use the model to understand, describe, and quantify important aspects of the physical system and predict the response of the system to inputs.

Throughout this text, we discuss models that allow for variations in the outputs of a system, even though the variables that we control are not purposely changed during our study. Figure 2.2 graphically displays a model that incorporates uncontrollable inputs (noise) that combine with the controllable inputs to produce the output of our system. Because of the uncontrollable inputs, the same settings for the controllable inputs do not result in identical outputs every time the system is measured.

Random Experiment

An experiment that can result in different outcomes, even though it is repeated in the same manner every time, is called a **random experiment**.

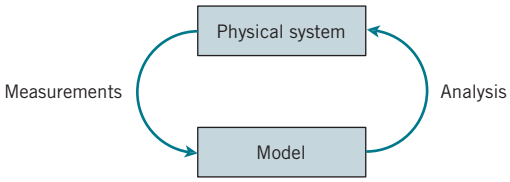


FIGURE 2.1

Continuous iteration between model and physical system.

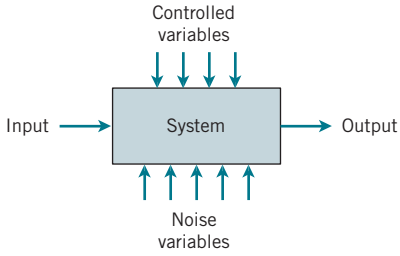


FIGURE 2.2

Noise variables affect the transformation of inputs to outputs.

For the example of measuring current in a copper wire, our model for the system might simply be Ohm's law. Because of uncontrollable inputs, variations in measurements of current are expected. Ohm's law might be a suitable approximation. However, if the variations are large relative to the intended use of the device under study, we might need to extend our model to include the variation. See Figure 2.3.

As another example, in the design of a communication system, such as a computer or voice communication network, the information capacity available to serve individuals using the network is an important design consideration. For voice communication, sufficient external lines need to be available to meet the requirements of a business. Assuming each line can carry only a single conversation, how many lines should be purchased? If too few lines are purchased, calls can be delayed or lost. The purchase of too many lines increases costs. Increasingly, design and product development is required to meet customer requirements *at a competitive cost*.

In the design of the voice communication system, a model is needed for the number of calls and the duration of calls. Even knowing that, on average, calls occur every five minutes and that they last five minutes is not sufficient. If calls arrived precisely at five-minute intervals and lasted for precisely five minutes, one phone line would be sufficient. However, the slightest variation in call number or duration would result in some calls being blocked by others. See Figure 2.4. A system designed without considering variation will be woefully inadequate for practical use. Our model for the number and duration of calls needs to include variation as an integral component.

2.1.2 Sample Spaces

To model and analyze a random experiment, we must understand the set of possible **outcomes** from the experiment. In this introduction to probability, we use the basic concepts of sets and operations on sets. It is assumed that the reader is familiar with these topics.

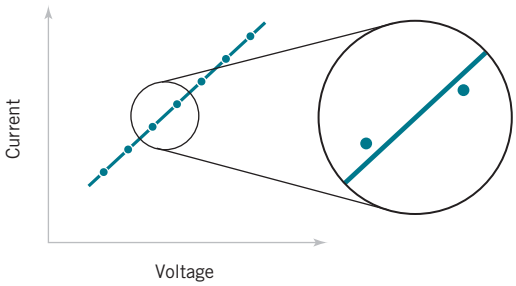


FIGURE 2.3

A closer examination of the system identifies deviations from the model.

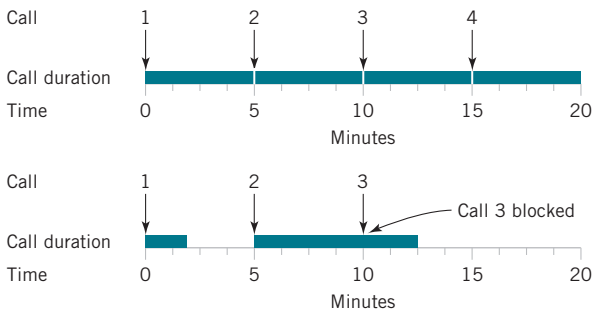


FIGURE 2.4

Variation causes disruptions in the system.

Sample Space

The set of all possible outcomes of a random experiment is called the **sample space** of the experiment. The sample space is denoted as S.

A sample space is often defined based on the objectives of the analysis. The following example illustrates several alternatives.

EXAMPLE 2.1 | Camera Flash

Consider an experiment that selects a cell phone camera and records the recycle time of a flash (the time taken to ready the camera for another flash). The possible values for this time depend on the resolution of the timer and on the minimum and maximum recycle times. However, because the time is positive, it is convenient to define the sample space as simply the positive real line

$$S = R^+ = \{x \mid x > 0\}$$

If it is known that all recycle times are between 1.5 and 5 seconds, the sample space can be

$$S = \{x \mid 1.5 < x < 5\}$$

If the objective of the analysis is to consider only whether the recycle time is low, medium, or high, the sample space can be taken to be the set of three outcomes

$$S = \{low, medium, high\}$$

If the objective is only to evaluate whether or not a particular camera conforms to a minimum recycle-time specification, the sample space can be simplified to a set of two outcomes

$$S = \{yes, no\}$$

that indicates whether or not the camera conforms.

It is useful to distinguish between two types of sample spaces.

Discrete and Continuous Sample Spaces

A sample space is **discrete** if it consists of a finite or countable infinite set of outcomes.

A sample space is **continuous** if it contains an interval (either finite or infinite) of real numbers.

In Example 2.1, the choice $S = R^+$ is an example of a continuous sample space, whereas $S = \{yes, no\}$ is a discrete sample space. As mentioned, the best choice of a sample space depends on the objectives of the study. As specific questions occur later in the book, appropriate sample spaces are discussed.

EXAMPLE 2.2 | Camera Specifications

Suppose that the recycle times of two cameras are recorded. The extension of the positive real line R is to take the sample space to be the positive quadrant of the plane

$$S = R^+ \times R^+$$

If the objective of the analysis is to consider only whether or not the cameras conform to the manufacturing specifications, either camera may or may not conform. We abbreviate *yes* and *no* as y and n. If the ordered pair yn indicates that the first camera conforms and the second does not, the sample space can be represented by the four outcomes:

$$S = \{yy, yn, ny, nn\}$$

If we are interested only in the number of conforming cameras in the sample, we might summarize the sample space as

$$S = \{0, 1, 2\}$$

As another example, consider an experiment in which cameras are tested unitl the flash recycle time fails to meet the specifications. The sample space can be represented as

$$S = \{n, yn, yyn, yyyn, yyyyn, \text{and so forth}\}$$

and this is an example of a discrete sample space that is countably infinite.

Sample spaces can also be described graphically with **tree diagrams**. When a sample space can be constructed in several steps or stages, we can represent each of the n_1 ways of completing the first step as a branch of a tree. Each of the ways of completing the second step can be represented as n_2 branches starting from the ends of the original branches, and so forth.

EXAMPLE 2.3 | Message Delays

Each message in a digital communication system is classified as to whether it is received within the time specified by the system design. If three messages are classified, use a tree diagram to represent the sample space of possible outcomes.

Each message can be received either on time or late. The possible results for three messages can be displayed by eight branches in the tree diagram shown in Figure 2.5.

Practical Interpretation: A tree diagram can effectively represent a sample space. Even if a tree becomes too large to construct, it can still conceptually clarify the sample space.

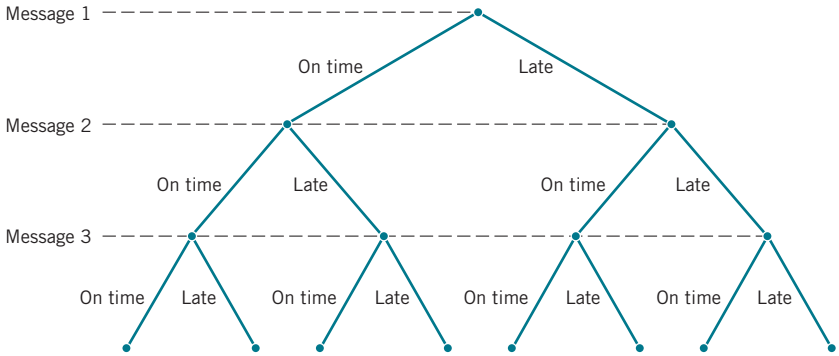


FIGURE 2.5
Tree diagram for three messages.

2.1.3 Events

Often we are interested in a collection of related outcomes from a random experiment. Related outcomes can be described by subsets of the sample space.

Event
An **event** is a subset of the sample space of a random experiment.

We can also be interested in describing new events from combinations of existing events. Because events are subsets, we can use basic set operations such as unions, intersections, and complements to form other events of interest. Some of the basic set operations are summarized here in terms of events:

- The *union* of two events is the event that consists of all outcomes that are contained in either of the two events. We denote the union as $E_1 \cup E_2$.
- The *intersection* of two events is the event that consists of all outcomes that are contained in both of the two events. We denote the intersection as $E_1 \cap E_2$.
- The *complement* of an event in a sample space is the set of outcomes in the sample space that are not in the event. We denote the complement of the event E as E'. The notation E^C is also used in other literature to denote the complement.

EXAMPLE 2.4 | Events

Consider the sample space $S = \{yy, yn, ny, nn\}$ in Example 2.2. Suppose that the subset of outcomes for which at least one camera conforms is denoted as E_1. Then,

$$E_1 = \{yy, yn, ny\}$$

The event such that both cameras do not conform, denoted as E_2, contains only the single outcome, $E_2 = \{nn\}$. Other examples of events are $E_3 = \emptyset$, the null set, and $E_4 = S$, the sample space. If $E_5 = \{yn, ny, nn\}$,

$$E_1 \cup E_5 = S \qquad E_1 \cap E_5 = \{yn, ny\} \qquad E_1' = \{nn\}$$

Practical Interpretation: Events are used to define outcomes of interest from a random experiment. One is often interested in the probabilities of specified events.

EXAMPLE 2.5 | Camera Recycle Time

As in Example 2.1, camera recycle times might use the sample space $S = R^+$, the set of positive real numbers. Let

$$E_1 = \{x \mid 10 \le x < 12\} \quad \text{and} \quad E_2 = \{x \mid 11 < x < 15\}$$

Then,

$$E_1 \cup E_2 = \{x \mid 10 \le x < 15\}$$

and

$$E_1 \cap E_2 = \{x \mid 11 < x < 12\}$$

Also,

$$E_1' = \{x \mid x < 10 \quad \text{or} \quad 12 \le x\}$$

and

$$E_1' \cap E_2 = \{x \mid 12 \le x < 15\}$$

EXAMPLE 2.6 | Hospital Emergency Visits

The following table summarizes visits to emergency departments at four hospitals in Arizona. People may leave without being seen by a physician, and those visits are denoted as LWBS. The remaining visits are serviced at the emergency department, and the visitor may or may not be admitted for a stay in the hospital.

Let A denote the event that a visit is to hospital 1, and let B denote the event that the result of the visit is LWBS. Calculate the number of outcomes in $A \cap B$, A', and $A \cup B$.

The event $A \cap B$ consists of the 195 visits to hospital 1 that result in LWBS. The event A' consists of the visits to hospitals 2, 3, and 4 and contains $6991 + 5640 + 4329 = 16{,}960$ visits.

The event $A \cup B$ consists of the visits to hospital 1 or the visits that result in LWBS, or both, and contains $5292 + 270 + 246 + 242 = 6050$ visits. Notice that the last result can also be calculated as the number of visits in A plus the number of visits in B minus the number of visits $A \cap B$ (that would otherwise be counted twice) $= 5292 + 953 - 195 = 6050$.

Practical Interpretation: Hospitals track visits that result in LWBS to understand resource needs and to improve patient services.

	Hospital				
	1	2	3	4	Total
Total	5292	6991	5640	4329	22,252
LWBS	195	270	246	242	953
Admitted	1277	1558	666	984	4485
Not admitted	3820	5163	4728	3103	16,814

Diagrams are often used to portray relationships between sets, and these diagrams are also used to describe relationships between events. We can use **Venn diagrams** to represent a sample space and events in a sample space. For example, in Figure 2.6(a) the sample space of the random experiment is represented as the points in the rectangle S. The events A and B are the subsets of points in the indicated regions. Figures 2.6(b) to 2.6(d) illustrate additional joint events. Figure 2.7 illustrates two events with no common outcomes.

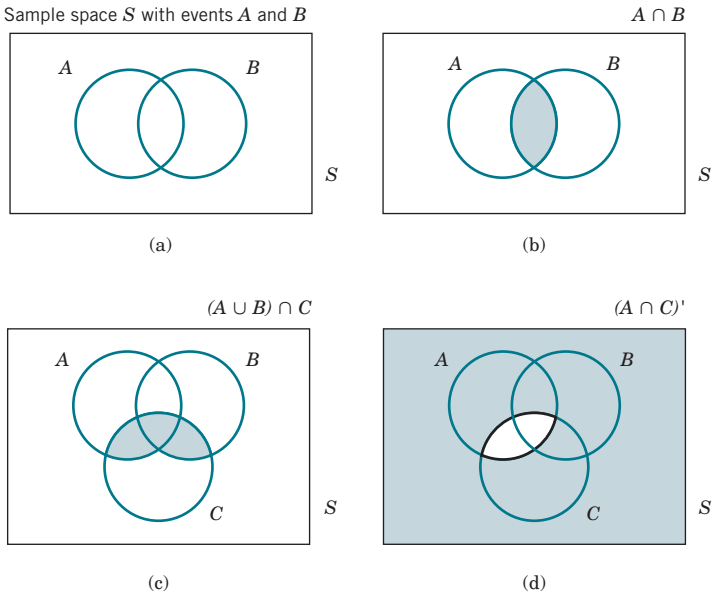


FIGURE 2.6
Venn diagrams.

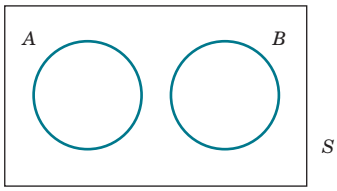


FIGURE 2.7
Mutually exclusive events.

Mutually Exclusive Events

Two events, denoted as E_1 and E_2, such that

$$E_1 \cap E_2 = \emptyset$$

are said to be **mutually exclusive**.

Additional results involving events are summarized in the following. The definition of the complement of an event implies that

$$(E')' = E$$

The distributive law for set operations implies that

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \quad \text{and} \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

DeMorgan's laws imply that

$$(A \cup B)' = A' \cap B' \quad \text{and} \quad (A \cap B)' = A' \cup B'$$

Also, remember that

$$A \cap B = B \cap A \quad \text{and} \quad A \cup B = B \cup A$$

2.2 Counting Techniques

In many of the examples in this chapter, it is easy to determine the number of outcomes in each event. In more complicated examples, determining the outcomes in the sample space (or an event) becomes more difficult. In these cases, counts of the numbers of outcomes in the sample space and various events are used to analyze the random experiments. These methods are referred to as **counting techniques**. Some simple rules can be used to simplify the calculations.

The tree diagram in Figure 2.5 describes the sample space for all possible arrival properties from three messages. The size of the sample space equals the number of branches in the last level of the tree, and this quantity equals $2 \times 2 \times 2 = 8$. This leads to the following useful result.

Multiplication Rule (for counting techniques)

Assume an operation can be described as a sequence of k steps, and

- the number of ways to complete step 1 is n_1, and
- the number of ways to complete step 2 is n_2 for each way to complete step 1, and
- the number of ways to complete step 3 is n_3 for each way to complete step 2, and
- so forth.

The total number of ways to complete the operation is

$$n_1 \times n_2 \times \cdots \times n_k$$

EXAMPLE 2.7 | Web Site Design

The design for a Website is to consist of four colors, three fonts, and three positions for an image. From the multiplication rule, $4 \times 3 \times 3 = 36$ different designs are possible.

Practical Interpretation: The use of the multiplication rule and other counting techniques enables one to easily determine the number of outcomes in a sample space or event and this, in turn, allows probabilities of events to be determined.

Permutations Another useful calculation finds the number of ordered sequences of the elements of a set. Consider a set of elements, such as $S = \{a, b, c\}$. A **permutation** of the elements is an ordered sequence of the elements. For example, *abc*, *acb*, *bac*, *bca*, *cab*, and *cba* are all of the permutations of the elements of S.

The number of **permutations** of n different elements is $n!$ where

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1 \tag{2.1}$$

This result follows from the multiplication rule. A permutation can be constructed as follows. Select the element to be placed in the first position of the sequence from the n elements, then select the element for the second position from the remaining $n - 1$ elements, then select the element for the third position from the remaining $n - 2$ elements, and so forth. Permutations such as these are sometimes referred to as *linear permutations*.

In some situations, we are interested in the number of arrangements of only some of the elements of a set. The following result also follows from the multiplication rule and the previous discussion.

Permutations of Subsets

The number of permutations of subsets of r elements selected from a set of n different elements is

$$P_r^n = n \times (n-1) \times (n-2) \times \cdots \times (n-r+1) = \frac{n!}{(n-r)!} \tag{2.2}$$

EXAMPLE 2.8 | Printed Circuit Board

A printed circuit board has eight different locations in which a component can be placed. If four different components are to be placed on the board, how many different designs are possible?

Each design consists of selecting a location from the eight locations for the first component, a location from the remaining seven for the second component, a location from the remaining six for the third component, and a location from the remaining five for the fourth component. Therefore,

$$P_4^8 = 8 \times 7 \times 6 \times 5 = \frac{8!}{4!}$$
$$= 1680 \text{ different designs are possible.}$$

Sometimes we are interested in counting the number of ordered sequences for objects that are not all different. The following result is a useful, general calculation.

Permutations of Similar Objects

The number of permutations of $n = n_1 + n_2 + \cdots + n_r$ objects of which n_1 are of one type, n_2 are of a second type, … , and n_r are of an rth type is

$$\frac{n!}{n_1!n_2!n_3! \ldots n_r!} \tag{2.3}$$

EXAMPLE 2.9 | Hospital Schedule

A hospital operating room needs to schedule three knee surgeries and two hip surgeries in a day. We denote a knee and hip surgery as k and h, respectively. The number of possible sequences of three knee and two hip surgeries is

$$\frac{5!}{2!3!} = 10$$

The 10 sequences are easily summarized:

$$\{kkkhh, kkhkh, kkhhk, khkkh, khkhk, khhkk, hkkkh, hkkhk, hkhkk, hhkkk\}$$

Combinations Another counting problem of interest is the number of subsets of r elements that can be selected from a set of n elements. Here, order is not important. These are called **combinations**. Every subset of r elements can be indicated by listing the elements in the set and marking each element with a "$*$" if it is to be included in the subset. Therefore, each permutation of $r*$'s and $n - r$ blanks indicates a different subset, and the numbers of these are obtained from Equation 2.3. For example, if the set is $S = \{a, b, c, d\}$, the subset $\{a, c\}$ can be indicated as

a	b	c	d
$*$		$*$	

Combinations

The number of combinations, subsets of r elements that can be selected from a set of n elements, is denoted as $\binom{n}{r}$ or C_r^n and

$$C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!} \tag{2.4}$$

EXAMPLE 2.10 | Printed Circuit Board Layout

A printed circuit board has eight different locations in which a component can be placed. If five identical components are to be placed on the board, how many different designs are possible?

Each design is a subset of size five from the eight locations that are to contain the components. From Equation 2.4, the number of possible designs is

$$\frac{8!}{5!3!} = 56$$

In random experiments in which items are selected from a batch, an item may or may not be replaced before the next one is selected. This is referred to as sampling **with** or **without replacement**, respectively. The following example uses the multiplication rule in combination with Equation 2.4 to answer a more difficult, but common, question for sampling without replacement.

EXAMPLE 2.11 | Sampling without Replacement

A bin of 50 manufactured parts contains 3 defective parts and 47 nondefective parts. A sample of 6 parts is selected without replacement. That is, each part can be selected only once, and the sample is a subset of the 50 parts. How many different samples are there of size 6 that contain exactly 2 defective parts?

A subset containing exactly 2 defective parts can be formed by first choosing the 2 defective parts from the 3 defective parts. Using Equation 2.4, the number of different ways this step can be completed is

$$\binom{3}{2} = \frac{3!}{2!1!} = 3$$

Then, the second step is to select the remaining 4 parts from the 47 acceptable parts in the bin. The number of different ways the second step can be completed is

$$\binom{47}{4} = \frac{47!}{4!43!} = 178{,}365$$

Therefore, from the multiplication rule, the number of subsets of size 6 that contain exactly 2 defective parts is

$$3 \times 178{,}365 = 535{,}095$$

As an additional computation, the total number of different subsets of size 6 is

$$\binom{50}{6} = \frac{50!}{6!44!} = 15{,}890{,}700$$

2.3 Interpretations and Axioms of Probability

In this chapter, we introduce probability for *discrete sample spaces*—those with only a finite (or countably infinite) set of outcomes. The restriction to these sample spaces enables us to simplify the concepts and the presentation without excessive mathematics.

Probability is used to quantify the likelihood, or chance, that an outcome of a random experiment will occur. "The chance of rain today is 30%" is a statement that quantifies our feeling about the possibility of rain. The likelihood of an outcome is quantified by assigning a number from the interval [0, 1] to the outcome (or a percentage from 0 to 100%). Higher numbers indicate that the outcome is more likely than lower numbers. A 0 indicates an outcome will not occur. A probability of 1 indicates that an outcome will occur with certainty.

The probability of an outcome can be interpreted as our subjective probability, or *degree of belief*, that the outcome will occur. Different individuals will no doubt assign different probabilities to the same outcomes. Another interpretation of probability is based on the conceptual model of repeated replications of the random experiment. The probability of an outcome is interpreted as the limiting value of the proportion of times the outcome occurs in n repetitions of the random experiment as n increases beyond all bounds. For example, if we assign probability 0.2 to the outcome that there is a corrupted pulse in a digital signal, we might interpret this assignment as implying that, if we analyze many pulses, approximately 20% of them will be corrupted. This

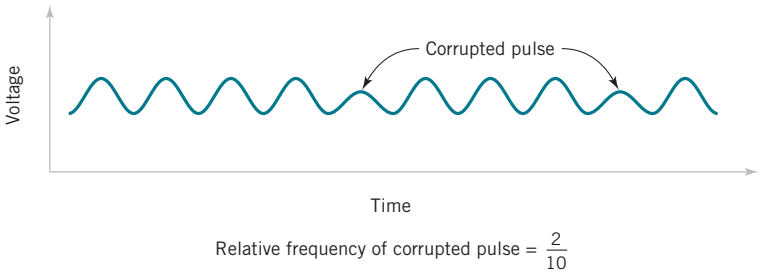


FIGURE 2.8

Relative frequency of corrupted pulses sent over a communication channel.

example provides a *relative frequency* interpretation of probability. The proportion, or relative frequency, of replications of the experiment that result in the outcome is 0.2. Probabilities are chosen so that the sum of the probabilities of all outcomes in an experiment adds up to 1. This convention facilitates the relative frequency interpretation of probability. Figure 2.8 illustrates the concept of relative frequency.

Probabilities for a random experiment are often assigned on the basis of a reasonable model of the system under study. One approach is to base probability assignments on the simple concept of equally likely outcomes. For example, suppose that we select 1 laser diode randomly from a batch of 100. *Randomly* implies that it is reasonable to assume that each diode in the batch has an equal chance of being selected. Because the sum of the probabilities must equal 1, the probability model for this experiment assigns probability of 0.01 to each of the 100 outcomes. We can interpret the probability by imagining many replications of the experiment. Each time we start with all 100 diodes and select 1 at random. The probability 0.01 assigned to a particular diode represents the proportion of replicates in which a particular diode is selected. When the model of **equally likely outcomes** is assumed, the probabilities are chosen to be equal.

Equally Likely Outcomes

Whenever a sample space consists of N possible outcomes that are equally likely, the probability of each outcome is $1/N$.

It is frequently necessary to assign probabilities to events that are composed of several outcomes from the sample space. This is straightforward for a discrete sample space.

EXAMPLE 2.12 | Laser Diodes

Assume that 30% of the laser diodes in a batch of 100 meet the minimum power requirements of a specific customer. If a laser diode is selected randomly, that is, each laser diode is equally likely to be selected, our intuitive feeling is that the probability of meeting the customer's requirements is 0.30.

Let E denote the subset of 30 diodes that meet the customer's requirements. Because E contains 30 outcomes and each outcome has probability 0.01, we conclude that the probability of E is 0.3. The conclusion matches our intuition. Figure 2.9 illustrates this example.

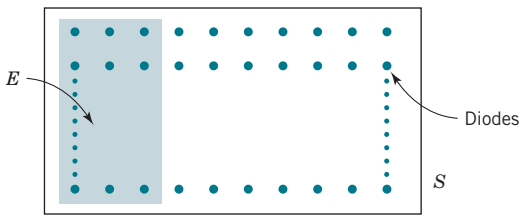


FIGURE 2.9

Probability of the event E is the sum of the probabilities of the outcomes in E.

For a discrete sample space, the probability of an event can be defined by the reasoning used in the preceding example.

Probability of an Event

For a discrete sample space, the *probability of an event* E, denoted as $P(E)$, equals the sum of the probabilities of the outcomes in E.

EXAMPLE 2.13 | Probabilities of Events

A random experiment can result in one of the outcomes $\{a, b, c, d\}$ with probabilities 0.1, 0.3, 0.5, and 0.1, respectively. Let A denote the event $\{a, b\}$, B the event $\{b, c, d\}$, and C the event $\{d\}$. Then,

$$P(A) = 0.1 + 0.3 = 0.4$$
$$P(B) = 0.3 + 0.5 + 0.1 = 0.9$$
$$P(C) = 0.1$$

Also, $P(A') = 0.6$, $P(B') = 0.1$, and $P(C') = 0.9$. Furthermore, because $A \cap B = \{b\}$, $P(A \cap B) = 0.3$. Because $A \cup B = \{a, b, c, d\}$, $P(A \cup B) = 0.1 + 0.3 + 0.5 + 0.1 = 1$. Because $A \cap C$ is the null set, $P(A \cap C) = 0$.

As another example, consider a random experiment in which more than one item is selected from a batch. In this case, *randomly* selected implies that each possible subset of items is equally likely.

EXAMPLE 2.14 | Manufacturing Inspection

Consider the inspection described in Example 2.11. From a bin of 50 parts, 6 parts are selected randomly without replacement. The bin contains 3 defective parts and 47 nondefective parts. What is the probability that exactly 2 defective parts are selected in the sample?

The sample space consists of all possible (unordered) subsets of 6 parts selected without replacement. As shown in Example 2.11, the number of subsets of size 6 that contain exactly 2 defective parts is 535,095 and the total number of subsets of size 6 is 15,890,700. The probability of an event is determined as the ratio of the number of outcomes in the event to the number of outcomes in the sample space (for equally likely outcomes). Therefore, the probability that a sample contains exactly 2 defective parts is

$$\frac{535{,}095}{15{,}890{,}700} = 0.034$$

A subset with no defective parts occurs when all 6 parts are selected from the 47 nondefective ones. Therefore, the number of subsets with no defective parts is

$$\frac{47!}{6!41!} = 10{,}737{,}573$$

and the probability that no defective parts are selected is

$$\frac{10{,}737{,}573}{15{,}890{,}700} = 0.676$$

Therefore, the sample of size 6 is likely to omit the defective parts. This example illustrates the hypergeometric distribution studied in Chapter 3.

Now that the probability of an event has been defined, we can collect the assumptions into a set of **axioms of probability** that must be satisfied in any random experiment. The axioms ensure that the probabilities assigned in an experiment can be interpreted as relative frequencies and that the assignments are consistent with our intuitive understanding of relationships between relative frequencies. For example, if event A is contained in event B, we should have $P(A) \leq P(B)$. The *axioms do not determine probabilities*; the probabilities are

assigned based on our knowledge of the system under study. However, the axioms enable us to easily calculate the probabilities of some events from knowledge of the probabilities of other events.

Axioms of Probability

Probability is a number that is assigned to each member of a collection of events from a random experiment that satisfies the following properties:

(1) $P(S) = 1$ where S is the sample space
(2) $0 \le P(E) \le 1$ for any event E
(3) For two events E_1 and E_2 with $E_1 \cap E_2 = \emptyset$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

The property that $0 \le P(E) \le 1$ is equivalent to the requirement that a relative frequency must be between 0 and 1. The property that $P(S) = 1$ is a consequence of the fact that an outcome from the sample space occurs on every trial of an experiment. Consequently, the relative frequency of S is 1. Property 3 implies that if the events E_1 and E_2 have no outcomes in common, the relative frequency of outcomes in $E_1 \cup E_2$ is the sum of the relative frequencies of the outcomes in E_1 and E_2.

These axioms imply the following results. The derivations are left as exercises. Now,

$$P(\emptyset) = 0$$

and for any event E,

$$P(E') = 1 - P(E)$$

For example, if the probability of the event E is 0.4, our interpretation of relative frequency implies that the probability of E' is 0.6. Furthermore, if the event E_1 is contained in the event E_2,

$$P(E_1) \le P(E_2)$$

2.4 Unions of Events and Addition Rules

Joint events are generated by applying basic set operations to individual events. Unions of events, such as $A \cup B$; intersections of events, such as $A \cap B$; and complements of events, such as A'—are commonly of interest. The probability of a joint event can often be determined from the probabilities of the individual events that it comprises. Basic set operations are also sometimes helpful in determining the probability of a joint event. In this section, the focus is on unions of events.

EXAMPLE 2.15 | Semiconductor Wafers

Table 2.1 lists the history of 940 wafers in a semiconductor manufacturing process. Suppose that 1 wafer is selected at random. Let H denote the event that the wafer contains high levels of contamination. Then, $P(H) = 358/940$.

Let C denote the event that the wafer is in the center of a sputtering tool. Then, $P(C) = 626/940$. Also, $P(H \cap C)$ is the probability that the wafer is from the center of the sputtering tool and contains high levels of contamination. Therefore,

$$P(H \cap C) = 112/940$$

The event $H \cup C$ is the event that a wafer is from the center of the sputtering tool or contains high levels of contamination (or both). From the table, $P(H \cup C) = 872/940$. An alternative

calculation of $P(H \cup C)$ can be obtained as follows. The 112 wafers in the event $H \cap C$ are included once in the calculation of $P(H)$ and again in the calculation of $P(C)$. Therefore, $P(H \cup C)$ can be determined to be

$$P(H \cup C) = P(H) + P(C) - P(H \cap C)$$
$$= \frac{358}{940} + \frac{626}{940} - \frac{112}{940} = \frac{872}{940}$$

Practical Interpretation: To better understand the sources of contamination, yield from different locations on wafers are routinely aggregated.

TABLE 2.1 Wafers in Semiconductor Manufacturing Classified by Contamination and Location

	Location in Sputtering Tool		
Contamination	Center	Edge	Total
Low	514	68	582
High	112	246	358
Total	626	314	

The preceding example illustrates that the probability of A or B is interpreted as $P(A \cup B)$ and that the following general **addition rule** applies.

Probability of a Union

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (2.5)$$

Recall that two events A and B are said to be mutually exclusive if $A \cap B = \emptyset$. Then, $P(A \cap B) = 0$, and the general result for the probability of $A \cup B$ simplifies to the third axiom of probability.

If A and B are mutually exclusive events,

$$P(A \cup B) = P(A) + P(B) \quad (2.6)$$

Three or More Events More complicated probabilities, such as $P(A \cup B \cup C)$, can be determined by repeated use of Equation 2.5 and by using some basic set operations. For example,

$$P(A \cup B \cup C) = P[(A \cup B) \cup C] = P(A \cup B) + P(C) - P[(A \cup B) \cap C]$$

Upon expanding $P(A \cup B)$ by Equation 2.5 and using the distributed rule for set operations to simplify $P[(A \cup B) \cap C]$, we obtain

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B)$$
$$- P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \quad (2.7)$$

Results for three or more events simplify considerably if the events are mutually exclusive. In general, a collection of events, $E_1, E_2, \ldots, E_k$, is said to be mutually exclusive if there is no overlap among any of them. The Venn diagram for several mutually exclusive events is shown in

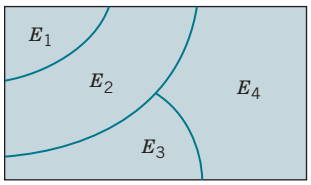


FIGURE 2.10
Venn diagram of four mutually exclusive events.

Figure 2.10. By generalizing the reasoning for the union of two events, the following result can be obtained:

Mutually Exclusive Events

A collection of events, $E_1, E_2, \ldots, E_k$, is said to be **mutually exclusive** if for all pairs,

$$E_i \cap E_j = \emptyset$$

For a collection of mutually exclusive events,

$$P(E_1 \cup E_2 \cup \ldots \cup E_k) = P(E_1) + P(E_2) + \cdots P(E_k) \tag{2.8}$$

EXAMPLE 2.16 | pH

Here is a simple example of mutually exclusive events, which are used quite frequently. Let X denote the pH of a sample. Consider the event that X is greater than 6.5 but less than or equal to 7.8. This probability is the sum of any collection of mutually exclusive events with union equal to the same range for X. One example is

$$P(6.5 < X \leq 7.8) = P(6.5 \leq X \leq 7.0) + P(7.0 < X \leq 7.5) + P(7.5 < X \leq 7.8)$$

Another example is

$$P(6.5 < X \leq 7.8) = P(6.5 < X \leq 6.6) + P(6.6 < X \leq 7.1) + P(7.1 < X \leq 7.4) + P(7.4 < X \leq 7.8)$$

The best choice depends on the particular probabilities available.

Practical Interpretation: The partition of an event into mutually exclusive subsets is widely used in later chapters to calculate probabilities.

2.5 Conditional Probability

Sometimes probabilities need to be reevaluated as additional information becomes available. A useful way to incorporate additional information into a probability model is to assume that the outcome that will be generated is a member of a given event. This event, say A, defines the conditions that the outcome will satisfy. Then probabilities can be revised to include this knowledge. The probability of an event B under the knowledge that the outcome will be in event A is denoted as

$$P(B \mid A)$$

and this is called the **conditional probability** of B given A.

A digital communication channel has an error rate of 1 bit per every 1000 transmitted. Errors are rare, but when they occur, they tend to occur in bursts that affect many consecutive bits. If a single bit is transmitted, we might model the probability of an error as 1/1000. However, if the

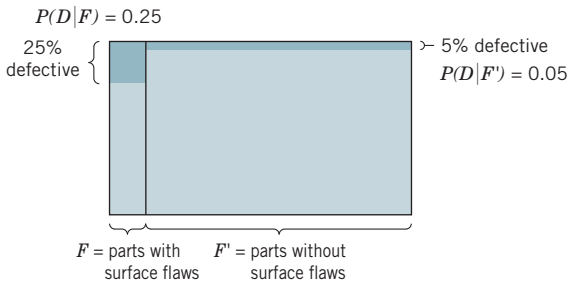


FIGURE 2.11

Conditional probabilities for parts with surface flaws.

previous bit was in error because of the bursts, we might believe that the probability that the next bit will be in error is greater than 1/1000.

In a thin film manufacturing process, the proportion of parts that are not acceptable is 2%. However, the process is sensitive to contamination problems that can increase the rate of parts that are not acceptable. If we knew that during a particular shift there were problems with the filters used to control contamination, we would assess the probability of a part being unacceptable as higher than 2%.

In a manufacturing process, 10% of the parts contain visible surface flaws and 25% of the parts with surface flaws are (functionally) defective parts. However, only 5% of parts without surface flaws are defective parts. The probability of a defective part depends on our knowledge of the presence or absence of a surface flaw. Let D denote the event that a part is defective, and let F denote the event that a part has a surface flaw. Then we denote the probability of D given or assuming that a part has a surface flaw, as $P(D \mid F)$. Because 25% of the parts with surface flaws are defective, our conclusion can be stated as $P(D \mid F) = 0.25$. Furthermore, because F' denotes the event that a part does not have a surface flaw and because 5% of the parts without surface flaws are defective, we have $P(D \mid F') = 0.05$. These results are shown graphically in Figure 2.11.

EXAMPLE 2.17 | Surface Flaws and Defectives

Table 2.2 provides an example of 400 parts classified by surface flaws and as (functionally) defective. For this table, the conditional probabilities match those discussed previously in this section. For example, of the parts with surface flaws (40 parts), the number of defective ones is 10. Therefore,

$$P(D \mid F) = 10/40 = 0.25$$

and of the parts without surface flaws (360 parts), the number of defective ones is 18. Therefore,

$$P(D \mid F') = 18/360 = 0.05$$

Practical Interpretation: The probability of being defective is five times greater for parts with surface flaws. This calculation illustrates how probabilities are adjusted for additional information. The result also suggests that there may be a link between surface flaws and functionally defective parts, which should be investigated.

TABLE 2.2 **Parts Classified**

		Surface Flaws		
		Yes (event F)	No	Total
Defective	Yes (event D)	10	18	28
	No	30	342	372
	Total	40	360	400

In Example 2.17, conditional probabilities were calculated directly. These probabilities can also be determined from the formal definition of conditional probability.

Conditional Probability

The **conditional probability** of an event B given an event A, denoted as $P(B \mid A)$, is

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \tag{2.9}$$

for $P(A) > 0$.

This definition can be understood in a special case in which all outcomes of a random experiment are equally likely. If there are N total outcomes,

$$P(A) = (\text{number of outcomes in } A)/N$$

Also,

$$P(A \cap B) = (\text{number of outcomes in } A \cap B)/N$$

Consequently,

$$\frac{P(A \cap B)}{P(A)} = \frac{\text{number of outcomes in } A \cap B}{\text{number of outcomes in } A}$$

Therefore, $P(B \mid A)$ can be interpreted as the relative frequency of event B among the trials that produce an outcome in event A.

EXAMPLE 2.18 | Tree Diagram

Again consider the 400 parts in Table 2.2. From this table,

$$P(D \mid F) = \frac{P(D \cap F)}{P(F)} = \frac{10}{400} \Big/ \frac{40}{400} = \frac{10}{40}$$

Note that in this example all four of the following probabilities are different:

$$P(F) = 40/400 \qquad P(F \mid D) = 10/28$$
$$P(D) = 28/400 \qquad P(D \mid F) = 10/40$$

Here, $P(D)$ and $P(D \mid F)$ are probabilities of the same event, but they are computed under two different states of knowledge. Similarly, $P(F)$ and $P(F \mid D)$ are computed under two different states of knowledge.

The tree diagram in Figure 2.12 can also be used to display conditional probabilities. The first branch is on surface flaw. Of the 40 parts with surface flaws, 10 are functionally defective and 30 are not. Therefore,

$$P(D \mid F) = 10/40 \quad \text{and} \quad P(D' \mid F) = 30/40$$

Of the 360 parts without surface flaws, 18 are functionally defective and 342 are not. Therefore,

$$P(D \mid F') = 18/360 \quad \text{and} \quad P(D' \mid F') = 342/360$$

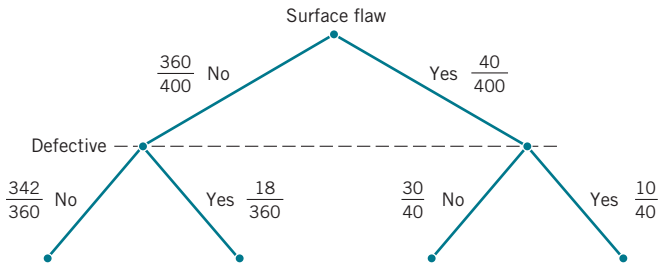


FIGURE 2.12

Tree diagram for parts classified.

Random Samples and Conditional Probability When sample spaces were presented earlier in this chapter, sampling with and without replacement were defined and illustrated for the simple case of a batch with three items $\{a, b, c\}$. If two items are selected randomly from this batch without replacement, each of the six outcomes in the ordered sample space $\{ab, ac, ba, bc, ca, cb\}$ has probability 1/6. If the unordered sample space is used, each of the three outcomes in $\{\{a, b\}, \{a, c\}, \{b, c\}\}$ has probability 1/3.

When a sample is selected randomly from a large batch, it is usually easier to avoid enumeration of the sample space and calculate probabilities from conditional probabilities. For example, suppose that a batch contains 10 parts from tool 1 and 40 parts from tool 2. If two parts are selected randomly, without replacement, what is the conditional probability that a part from tool 2 is selected second given that a part from tool 1 is selected first?

Although the answer can be determined from counts of outcomes, this type of question can be answered more easily with the following result.

Random Samples

To select *randomly* implies that at each step of the sample, the items that remain in the batch are equally likely to be selected.

If a part from tool 1 were selected with the first pick, 49 items would remain, 9 from tool 1 and 40 from tool 2, and they would be equally likely to be picked. Therefore, the probability that a part from tool 2 would be selected with the second pick given this first pick is

$$P(E_2 \mid E_1) = 40/49$$

In this manner, other probabilities can also be simplified. For example, let the event E consist of the outcomes with the first selected part from tool 1 and the second part from tool 2. To determine the probability of E, consider each step. The probability that a part from tool 1 is selected with the first pick is $P(E_1) = 10/50$. The conditional probability that a part from tool 2 is selected with the second pick, given that a part from tool 1 is selected first, is $P(E_2 \mid E_1) = 40/49$. Therefore,

$$P(E) = P(E_2 \mid E_1)\, P(E_1) = \frac{40}{49} \cdot \frac{10}{50} = \frac{8}{49}$$

Sometimes a partition of the question into successive picks is an easier method to solve the problem.

2.6 Intersections of Events and Multiplication and Total Probability Rules

The probability of the intersection of two events is often needed. The conditional probability definition in Equation 2.9 can be rewritten to provide a formula known as the **multiplication rule** for probabilities.

Multiplication Rule

$$P(A \cap B) = P(B \mid A)P(A) = P(A \mid B)P(B) \tag{2.10}$$

The last expression in Equation 2.10 is obtained by interchanging A and B.

EXAMPLE 2.19 | Machining Stages

The probability that the first stage of a numerically controlled machining operation for high-rpm pistons meets specifications is 0.90. Failures are due to metal variations, fixture alignment, cutting blade condition, vibration, and ambient environmental conditions. Given that the first stage meets specifications, the probability that a second stage of machining meets specifications is 0.95. What is the probability that both stages meet specifications?

Let A and B denote the events that the first and second stages meet specifications, respectively. The probability requested is

$$P(A \cap B) = P(B \mid A)P(A) = 0.95(0.90) = 0.855$$

Although it is also true that $P(A \cap B) = P(A \mid B)P(B)$, the information provided in the problem does not match this second formulation.

Practical Interpretation: The probability that both stages meet specifications is approximately 0.85, and if additional stages were needed to complete a piston, the probability would decrease further. Consequently, the probability that each stage is completed successfully needs to be large in order for a piston to meet all specifications.

Sometimes the probability of an event is given under each of several conditions. With enough of these conditional probabilities, the probability of the event can be recovered. For example, suppose that in semiconductor manufacturing, the probability is 0.10 that a chip subjected to high levels of contamination during manufacturing causes a product failure. The probability is 0.005 that a chip not subjected to high contamination levels during manufacturing causes a product failure. In a particular production run, 20% of the chips are subject to high levels of contamination. What is the probability that a product using one of these chips fails?

Clearly, the requested probability depends on whether or not the chip was exposed to high levels of contamination. For any event B, we can write B as the union of the part of B in A and the part of B in A'. That is,

$$B = (A \cap B) \cup (A' \cap B)$$

This result is shown in the Venn diagram in Figure 2.13. Because A and A' are mutually exclusive, $A \cap B$ and $A' \cap B$ are mutually exclusive. Therefore, from the probability of the union of mutually exclusive events in Equation 2.6 and the multiplication rule in Equation 2.10, the following **total probability rule** is obtained.

Total Probability Rule (Two Events)

For any events A and B,

$$P(B) = P(B \cap A) + P(B \cap A') = P(B \mid A)P(A) + P(B \mid A')P(A') \qquad (2.11)$$

EXAMPLE 2.20 | Semiconductor Contamination

Consider the contamination discussion at the start of this section. The information is summarized here.

Probability of Failure	Level of Contamination	Probability of Level
0.1	High	0.2
0.005	Not high	0.8

Let F denote the event that the product fails, and let H denote the event that the chip is exposed to high levels of contamination. The requested probability is $P(F)$, and the information provided can be represented as

$$P(F \mid H) = 0.10 \quad \text{and} \quad P(F \mid H') = 0.005$$
$$P(H) = 0.20 \quad \text{and} \quad P(H') = 0.80$$

From Equation 2.11,

$$P(F) = 0.10(0.20) + 0.005(0.80) = 0.024$$

which can be interpreted as just the weighted average of the two probabilities of failure.

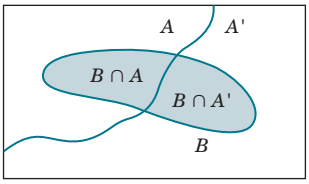
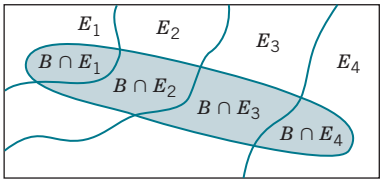


FIGURE 2.13
Partitioning an event into two mutually exclusive subsets.



$B = (B \cap E_1) \cup (B \cap E_2) \cup (B \cap E_3) \cup (B \cap E_4)$

FIGURE 2.14
Partitioning an event into several mutually exclusive subsets.

The reasoning used to develop Equation 2.11 can be applied more generally. Because $A \cup A' = S$, we know $(A \cap B) \cup (A' \cap B)$ equals B, and because $A \cap A' = \phi$, we know $A \cap B$ and $A' \cap B$ are mutually exclusive. In general, a collection of sets $E_1, E_2, \ldots, E_k$ such that $E_1 \cup E_2 \cup \ldots \cup E_k = S$ is said to be *exhaustive*. A graphical display of partitioning an event B among a collection of mutually exclusive and exhaustive events is shown in Figure 2.14.

Total Probability Rule (Multiple Events)

Assume $E_1, E_2, \ldots, E_k$ are k mutually exclusive and exhaustive sets. Then

$$\begin{aligned} P(B) &= P(B \cap E_1) + P(B \cap E_2) + \cdots + P(B \cap E_k) \\ &= P(B \mid E_1)P(E_1) + P(B \mid E_2)P(E_2) + \cdots + P(B \mid E_k)P(E_k) \end{aligned} \quad (2.12)$$

2.7 Independence

In some cases, the conditional probability of $P(B \mid A)$ might equal $P(B)$. In this special case, knowledge that the outcome of the experiment is in event A does not affect the probability that the outcome is in event B.

EXAMPLE 2.21 | Sampling with Replacement

Consider the inspection described in Example 2.11. Six parts are selected randomly from a bin of 50 parts, but assume that the selected part is replaced before the next one is selected. The bin contains 3 defective parts and 47 nondefective parts. What is the probability that the second part is defective given that the first part is defective?

In shorthand notation, the requested probability is $P(B \mid A)$, where A and B denote the events that the first and second parts are defective, respectively. Because the first part is replaced prior to selecting the second part, the bin still contains 50 parts, of which 3 are defective. Therefore, the probability of B does not depend on whether or not the first part is defective. That is,

$$P(B \mid A) = \frac{3}{50}$$

Also, the probability that both parts are defective is

$$P(A \cap B) = P(B \mid A)P(A) = \frac{3}{50} \cdot \frac{3}{50} = \frac{9}{2500}$$

The preceding example illustrates the following conclusions. In the special case that $P(B \mid A) = P(B)$, we obtain

$$P(A \cap B) = P(B \mid A)P(A) = P(B)P(A)$$

and

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

These conclusions lead to an important definition.

Independence (two events)

Two events are **independent** if any one of the following equivalent statements is true:

(1) $P(A \mid B) = P(A)$

(2) $P(B \mid A) = P(B)$

(3) $P(A \cap B) = P(A)P(B)$ (2.13)

It is left as an exercise to show that independence implies related results such as

$$P(A' \cap B') = P(A')P(B')$$

The concept of independence is an important relationship between events and is used throughout this text. A mutually exclusive relationship between two events is based only on the outcomes that compose the events. However, an independence relationship depends on the probability model used for the random experiment. Often, independence is assumed to be part of the random experiment that describes the physical system under study.

EXAMPLE 2.22

Consider the inspection described in Example 2.11. Six parts are selected randomly without replacement from a bin of 50 parts. The bin contains 3 defective parts and 47 nondefective parts. Let A and B denote the events that the first and second parts are defective, respectively.

We suspect that these two events are not independent because the knowledge that the first part is defective suggests that it is less likely that the second part selected is defective. Indeed, $P(B \mid A) = 2/49$. Now, what is $P(B)$? Finding the unconditional $P(B)$ takes some work because the possible values of the first selection need to be considered:

$$P(B) = P(B \mid A)P(A) + P(B \mid A')P(A')$$
$$= \frac{2}{49} \cdot \frac{3}{50} + \frac{3}{49} \cdot \frac{47}{50} = \frac{3}{50}$$

Interestingly, $P(B)$, the unconditional probability that the second part selected is defective, without any knowledge of the first part, is the same as the probability that the first part selected is defective. Yet our goal is to assess independence. Because $P(B \mid A)$ does not equal $P(B)$, the two events are not independent, as we expected.

When considering three or more events, we can extend the definition of independence with the following general result.

Independence (multiple events)

The events $E_1, E_2, \ldots, E_n$ are independent if and only if for any subset of these events

$$P\left(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}\right) = P\left(E_{i_1}\right) \times P\left(E_{i_2}\right) \times \cdots \times P\left(E_{i_k}\right) \quad (2.14)$$

This definition is typically used to calculate the probability that several events occur, assuming that they are independent and the individual event probabilities are known. The knowledge that the events are independent usually comes from a fundamental understanding of the random experiment.

EXAMPLE 2.23 | Series Circuit

The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently. What is the probability that the circuit operates?

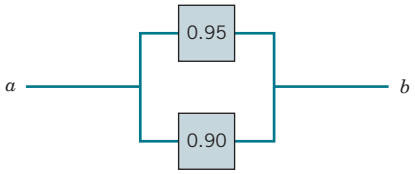
Let L and R denote the events that the left and right devices operate, respectively. There is a path only if both operate. The probability that the circuit operates is

$$P(L \text{ and } R) = P(L \cap R) = P(L)P(R) = 0.80(0.90) = 0.72$$

Practical Interpretation: Notice that the probability that the circuit operates degrades to approximately 0.7 when all devices are required to be functional. The probability that each device is functional needs to be large for a circuit to operate when many devices are connected in series.

EXAMPLE 2.24 | Parallel Circuit

The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently. What is the probability that the circuit operates?



Let T and B denote the events that the top and bottom devices operate, respectively. There is a path if at least one device operates. The probability that the circuit operates is

$$P(T \text{ or } B) = 1 - P[(T \text{ or } B)'] = 1 - P(T' \text{and } B')$$

A simple formula for the solution can be derived from the complements T' and B'. From the independence assumption,

$$P(T' \text{and } B') = P(T')P(B') = (1 - 0.95)(1 - 0.90) = 0.005$$

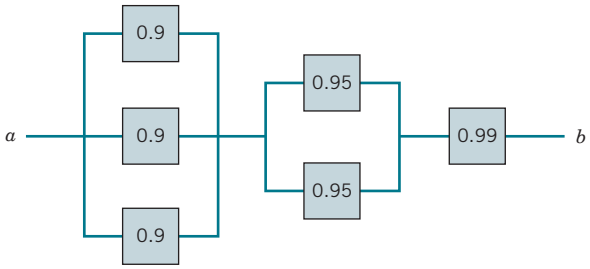
so

$$P(T \text{ or } B) = 1 - 0.005 = 0.995$$

Practical Interpretation: Notice that the probability that the circuit operates is larger than the probability that either device is functional. This is an advantage of a parallel architecture. A disadvantage is that multiple devices are needed.

EXAMPLE 2.25 | Advanced Circuit

The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently. What is the probability that the circuit operates?



The solution can be obtained from a partition of the graph into three columns. Let L denote the event that there is a path of functional devices only through the three units on the left. From the independence and based on the previous example,

$$P(L) = 1 - 0.1^3$$

Similarly, let M denote the event that there is a path of functional devices only through the two units in the middle. Then,

$$P(M) = 1 - 0.05^2$$

The probability that there is a path of functional devices only through the one unit on the right is simply the probability that the device functions, namely, 0.99. Therefore, with the independence assumption used again, the solution is

$$(1 - 0.1^3)(1 - 0.05^2)(0.99) = 0.987$$

2.8 Bayes' Theorem

The examples in this chapter indicate that information is often presented in terms of conditional probabilities. These conditional probabilities commonly provide the probability of an event (such as failure) given a condition (such as high or low contamination). But after a random experiment generates an outcome, we are naturally interested in the probability that a condition was present (high contamination) given an outcome (a semiconductor failure). Thomas Bayes addressed this essential question in the 1700s and developed the fundamental result known as **Bayes' theorem**. Do not let the simplicity of the mathematics conceal the importance. There is extensive interest in such probabilities in modern statistical analysis.

From the definition of conditional probability,

$$P(A \cap B) = P(A \mid B)P(B) = P(B \cap A) = P(B \mid A)P(A)$$

Now, considering the second and last terms in the preceding expression, we can write

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \quad \text{for} \quad P(B) > 0 \tag{2.15}$$

This is a useful result that enables us to solve for $P(A \mid B)$ in terms of $P(B \mid A)$.

EXAMPLE 2.26

Reconsider Example 2.20. The conditional probability that a high level of contamination was present when a failure occurred is to be determined. The information from Example 2.20 is summarized here.

Probability of Failure	Level of Contamination	Probability of Level
0.1	High	0.2
0.005	Not high	0.8

The probability of $P(H \mid F)$ is determined from

$$P(H \mid F) = \frac{P(F \mid H)P(H)}{P(F)} = \frac{0.10(0.20)}{0.024} = 0.83$$

The value of $P(F)$ in the denominator of our solution was found from $P(F) = P(F \mid H)P(H) + P(F \mid H')P(H')$.

In general, if $P(B)$ in the denominator of Equation 2.15 is written using the total probability rule in Equation 2.12, we obtain the following general result, which is known as **Bayes' theorem**.

Bayes' Theorem

If $E_1, E_2, \ldots, E_k$ are k mutually exclusive and exhaustive events and B is any event,

$$P(E_1 \mid B) = \frac{P(B \mid E_1)P(E_1)}{P(B \mid E_1)P(E_1) + P(B \mid E_2)P(E_2) + \cdots + P(B \mid E_k)P(E_k)} \tag{2.16}$$

for $P(B) > 0$

Notice that the numerator always equals one of the terms in the sum in the denominator.

EXAMPLE 2.27 | Medical Diagnostic

Because a new medical procedure has been shown to be effective in the early detection of an illness, a medical screening of the population is proposed. The probability that the test correctly identifies someone with the illness as positive (known as the sensitivity) is 0.99, and the probability that the test correctly identifies someone without the illness as negative (known as the specificity) is 0.95. The incidence of the illness in the general population is 0.0001. You take the test, and the result is positive. What is the probability that you have the illness?

Let D denote the event that you have the illness, and let S denote the event that the test signals positive. The probability requested can be denoted as $P(D \mid S)$. The probability that the test correctly signals someone without the illness as negative is 0.95. Consequently, the probability of a positive test without the illness is

$$P(S \mid D') = 0.05$$

From Bayes' theorem,

$$\begin{aligned} P(D \mid S) &= \frac{P(S \mid D)P(D)}{P(S \mid D)P(D) + P(S \mid D')P(D')} \\ &= \frac{0.99(0.0001)}{0.99(0.0001) + 0.05(1 - 0.0001)} \\ &= 1/506 = 0.002 \end{aligned}$$

Practical Interpretation: The probability that the illness is present given a positive result from the test is only 0.002. Surprisingly, even though the test is effective, in the sense that the sensitivity and specificity are both high, because the incidence of the illness in the general population is low, the chance is quite small that you actually have the disease even with a positive result.

2.9 Random Variables

We often summarize the outcome from a random experiment by a simple number. In many of the examples of random experiments that we have considered, the sample space has been a description of possible outcomes. In some cases, descriptions of outcomes are sufficient, but in other cases, it is useful to associate a number with each outcome in the sample space. Because the particular outcome of the experiment is not known in advance, the resulting value of our variable is not known in advance. For this reason, the variable that associates a number with the outcome of a random experiment is referred to as a **random variable**.

Random Variable

A **random variable** is a function that assigns a real number to each outcome in the sample space of a random experiment.

Notation is used to distinguish between a random variable and the real number.

Notation

A random variable is denoted by an uppercase letter such as X. After an experiment is conducted, the measured value of the random variable is denoted by a lowercase letter such as $x = 70$ milliamperes.

Sometimes a measurement (such as current in a copper wire or length of a machined part) can assume any value in an interval of real numbers (at least theoretically). Then arbitrary precision in the measurement is possible. Of course, in practice, we might round off to the nearest tenth

or hundredth of a unit. The random variable that represents this measurement is said to be a **continuous random variable**. The range of the random variable includes all values in an interval of real numbers; that is, the range can be thought of as a continuum.

In other experiments, we might record a count such as the number of transmitted bits that are received in error. Then, the measurement is limited to integers. Or we might record that a proportion such as 0.0042 of the 10,000 transmitted bits were received in error. Then, the measurement is fractional, but it is still limited to discrete points on the real line. Whenever the measurement is limited to discrete points on the real line, the random variable is said to be a **discrete random variable**.

Discrete and Continuous Random Variables

A **discrete random variable** is a random variable with a finite (or countably infinite) range.

A **continuous random variable** is a random variable with an interval (either finite or infinite) of real numbers for its range.

In some cases, the random variable X is actually discrete but, because the range of possible values is so large, it might be more convenient to analyze X as a continuous random variable. For example, suppose that current measurements are read from a digital instrument that displays the current to the nearest 100th of a milliampere. Because the possible measurements are limited, the random variable is discrete. However, it might be a more convenient, simple approximation to assume that the current measurements are values of a continuous random variable.

Examples of Random Variables

Examples of **continuous** random variables:

electrical current, length, pressure, temperature, time, voltage, weight

Examples of **discrete** random variables:

number of scratches on a surface, proportion of defective parts among 1000 tested, number of transmitted bits received in error

Important Terms and Concepts

Addition rule
Axioms of probability
Bayes' theorem
Combination
Conditional probability
Counting techniques
Equally likely outcomes
Event
Independence
Multiplication rule
Mutually exclusive events
Outcome
Permutation
Probability
Random samples
Random variables—discrete and continuous
Sample spaces—discrete and continuous
Total probability rule
Tree diagram
Venn diagram
With or without replacement

CHAPTER 3

Discrete Random Variables and Probability Distributions

Thinkstock Images/Getty Images

LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Determine probabilities from probability mass functions and the reverse
2. Determine probabilities and probability mass functions from cumulative distribution functions and the reverse
3. Calculate means and variances for discrete random variables
4. Understand the assumptions for some common discrete probability distributions
5. Calculate probabilities and determine means and variances for some common discrete probability distributions

CHAPTER OUTLINE

3.1 Probability Distributions and Probability Mass Functions

3.2 Cumulative Distribution Functions

3.3 Mean and Variance of a Discrete Random Variable

3.4 Discrete Uniform Distribution

3.5 Binomial Distribution

3.6 Geometric and Negative Binomial Distributions

3.7 Hypergeometric Distribution

3.8 Poisson Distribution

A redundant array of independent disks (RAID) uses multiple physical disk drives as one logical unit in a computer system. The array can increase performance and robustness to a disk failure. Data copies can be written simultaneously to multiple drives (known as *mirroring*) to provide immediate backup and the ability to recover from failures but with less storage capacity than would otherwise be available. Alternatively, to increase performance, the data can be distributed among multiple disks with only a fraction of the data on each one (known as *striping*). But a failure to even a single disk can lead to loss of data. An intermediate design is to distribute the source data along with additional data (known as *parity data*) across multiple disks. With the parity data, the source data can be recovered even with disk failures. In particular, a RAID 5 design uses striping and parity to be able to recover the source data if one disk in the array fails, and a RAID 6 design allows for data recovery even if two disks fail. Disk failures due to hardware malfunction are often assumed to be independent with constant probability. With a large number of disks in an array, the risk of data loss and the appropriate array design to meet the system performance, availability, and cost criteria are important. The number of failed drives can be modeled as a discrete random variable; the risk of data loss in a redundant system is only one example of the use of the topics in this chapter.

3.1 Probability Distributions and Probability Mass Functions

Many physical systems can be modeled by the same or similar random experiments and random variables. The distribution of the random variables involved in each of these common systems can be analyzed, and the results can be used in different applications and examples. In this chapter, we present the analysis of several random experiments and **discrete random variables** that frequently arise in applications.

Random variables are so important in random experiments that sometimes we essentially ignore the original sample space of the experiment and focus on the probability distribution of the random variable. In Example 3.1, we might summarize the random experiment in terms of the three possible values of X, namely $\{0, 1, 2\}$. In this manner, a random variable can simplify the description and analysis of a random experiment.

The **probability distribution** of a random variable X is a description of the probabilities associated with the possible values of X. For a discrete random variable, the distribution is often specified by just a list of the possible values along with the probability of each. For other cases, probabilities are expressed in terms of a formula.

EXAMPLE 3.1 | Flash Recharge Time

The time to recharge the flash is tested in three cell-phone cameras. The probability that a camera meets the recharge specification is 0.8, and the cameras perform independently. See Table 3.1 for the sample space for the experiment and associated probabilities. For example, because the cameras are independent, the probability that the first and second cameras pass the test and the third one fails, denoted as ppf, is

$$P(ppf) = (0.8)(0.8)(0.2) = 0.128$$

The random variable X denotes the number of cameras that pass the test. The last column of the table shows the values of X assigned to each outcome of the experiment.

TABLE 3.1 Camera Flash Tests

Camera 1	Camera 2	Camera 3	Probability	X
Pass	Pass	Pass	0.512	3
Fail	Pass	Pass	0.128	2
Pass	Fail	Pass	0.128	2
Fail	Fail	Pass	0.032	1
Pass	Pass	Fail	0.128	2
Fail	Pass	Fail	0.032	1
Pass	Fail	Fail	0.032	1
Fail	Fail	Fail	0.008	0

EXAMPLE 3.2 | Particles of Contamination

Define the random variable X to be the number of contamination particles on a wafer in semiconductor manufacturing. Although wafers possess a number of characteristics, the random variable X summarizes the wafer only in terms of the number of particles.

The possible values of X are integers from zero up to some large value that represents the maximum number of particles that can be found on one of the wafers. If this maximum number is large, we might simply assume that the range of X is the set of integers from zero to infinity.

Note that more than one random variable can be defined in an experiment. We might also define the random variable Y to be the number of chips from a wafer that fails the final test.

EXAMPLE 3.3 | Digital Channel

There is a chance that a bit transmitted through a digital transmission channel is received in error. Let X equal the number of bits in error in the next four bits transmitted. The possible values for X are $\{0, 1, 2, 3, 4\}$. Based on a model for the errors that is presented in the following section, probabilities for these values will be determined. Suppose that the probabilities are

$$P(X = 0) = 0.6561 \quad P(X = 1) = 0.2916$$
$$P(X = 2) = 0.0486 \quad P(X = 3) = 0.0036$$
$$P(X = 4) = 0.0001$$

The probability distribution of X is specified by the possible values along with the probability of each. A graphical description of the probability distribution of X is shown in Figure 3.1.

Practical Interpretation: A random experiment can often be summarized with a random variable and its distribution. The details of the sample space can often be omitted.

Suppose that a loading on a long, thin beam places mass only at discrete points. See Figure 3.2. The loading can be described by a function that specifies the mass at each of the discrete points. Similarly, for a discrete random variable X, its distribution can be described by a function that specifies the probability at each of the possible discrete values for X.

Probability Mass Function

For a discrete random variable X with possible values $x_1, x_2, \ldots, x_n$, a **probability mass function** is a function such that

$$(1)\ f(x_i) \geq 0$$
$$(2)\ \sum_{i=1}^{n} f(x_i) = 1$$
$$(3)\ f(x_i) = P(X = x_i) \qquad (3.1)$$

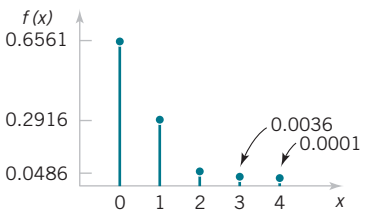


FIGURE 3.1
Probability distribution for bits in error.

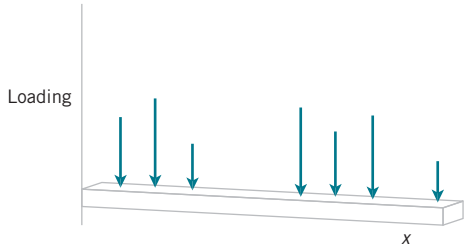


FIGURE 3.2
Loadings at discrete points on a long, thin beam.

For the bits in error in Example 3.3, $f(0) = 0.6561, f(1) = 0.2916, f(2) = 0.0486, f(3) = 0.0036$, and $f(4) = 0.0001$. Check that the probabilities sum to 1.

EXAMPLE 3.4 | Wafer Contamination

Let the random variable X denote the number of semiconductor wafers that need to be analyzed in order to detect a large particle of contamination. Assume that the probability that a wafer contains a large particle is 0.01 and that the wafers are independent. Determine the probability distribution of X.

Let p denote a wafer in which a large particle is present, and let a denote a wafer in which it is absent. The sample space of the experiment is infinite, and it can be represented as all possible sequences that start with a string of a's and end with p. That is,

$$s = \{p, ap, aap, aaap, aaaap, aaaaap, \text{and so forth}\}$$

Consider a few special cases. We have $P(X = 1) = P(p) = 0.01$. Also, using the independence assumption,

$$P(X = 2) = P(ap) = 0.99(0.01) = 0.0099$$

A general formula is

$$P(X = x) = P(\underbrace{aa \ldots ap}_{(x-1)a\text{'s}}) = 0.99^{x-1}(0.01), \text{ for } x = 1, 2, 3, \ldots$$

Describing the probabilities associated with X in terms of this formula is a simple method to define the distribution of X in this example. Clearly $f(x) \geq 0$. The fact that the sum of the probabilities is 1 is left as an exercise. This is an example of a geometric random variable for which details are provided later in this chapter.

Practical Interpretation: The random experiment here has an unbounded number of outcomes, but it can still be conveniently modeled with a discrete random variable with a (countably) infinite range.

3.2 Cumulative Distribution Functions

An alternate method for describing a random variable's probability distribution is with cumulative probabilities such as $P(X \leq x)$. Furthermore, cumulative probabilities can be used to find the probability mass function of a discrete random variable. Consider the following example.

EXAMPLE 3.5 | Digital Channel

In Example 3.3, we might be interested in the probability that three or fewer bits are in error. This question can be expressed as $P(X \leq 3)$.

The event that $\{X \leq 3\}$ is the union of the events $\{X = 0\}, \{X = 1\}, \{X = 2\}$, and $\{X = 3\}$. Clearly, these three events are mutually exclusive. Therefore,

$$\begin{aligned} P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= 0.6561 + 0.2916 + 0.0486 + 0.0036 = 0.9999 \end{aligned}$$

This approach can also be used to determine

$$P(X = 3) = P(X \leq 3) - P(X \leq 2) = 0.0036$$

In general, for any discrete random variable with possible values $x_1, x_2, \ldots$, the events $\{X = x_1\}, \{X = x_2\}, \ldots$ are mutually exclusive. Therefore, $P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$. This leads to the following definition.

Cumulative Distribution Function

The **cumulative distribution function** of a discrete random variable X, denoted as $F(x)$, is

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

For a discrete random variable X, $F(x)$ satisfies the following properties.

$$(1)\ F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$$
$$(2)\ 0 \le F(x) \le 1$$
$$(3)\ \text{If } x \le y, \text{ then } F(x) \le F(y) \qquad (3.2)$$

Properties (1) and (2) of a cumulative distribution function follow from the definition. Property (3) follows from the fact that if $x \le y$, the event that $\{X \le x\}$ is contained in the event $\{X \le y\}$. Like a probability mass function, a cumulative distribution function provides probabilities.

Even if the random variable X can assume only integer values, the cumulative distribution function is defined at noninteger values. In Example 3.5, $F(1.5) = P(X \le 1.5) = P\{X = 0\} + P(X = 1) = 0.6561 + 0.2916 = 0.9477$. Also, $F(x) = 0.9477$ for all $1 \le x < 2$ and

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.6561 & 0 \le x < 1 \\ 0.9477 & 1 \le x < 2 \\ 0.9963 & 2 \le x < 3 \\ 0.9999 & 3 \le x < 4 \\ 1 & 4 \le x \end{cases}$$

That is, $F(x)$ is piecewise constant between the values $x_1, x_2, \ldots$.

Furthermore, $P(X = x_i)$ can be determined from the *jump* at the value x_i. More specifically,

$$P(X = x_i) = F(x_i) - \lim_{x \uparrow x_i} F(x)$$

and this expression calculates the difference between $F(x_i)$ and the limit of $F(x)$ as x increases to x_i.

EXAMPLE 3.6 | Cumulative Distribution Function

Determine the probability mass function of X from the following cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < -2 \\ 0.2 & -2 \le x < 0 \\ 0.7 & 0 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

Figure 3.3 displays a plot of $F(x)$. From the plot, the only points that receive nonzero probability are -2, 0, and 2. The probability mass function at each point is the jump in the cumulative distribution function at the point. Therefore,

$$f(-2) = 0.2 - 0 = 0.2$$
$$f(0) = 0.7 - 0.2 = 0.5$$
$$f(2) = 1.0 - 0.7 = 0.3$$

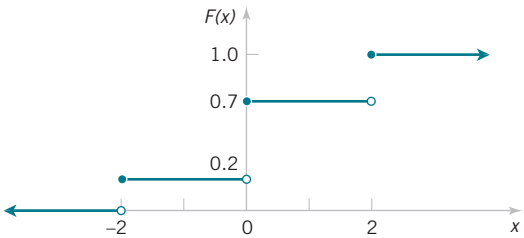


FIGURE 3.3
Cumulative distribution function for Example 3.6.

3.3 Mean and Variance of a Discrete Random Variable

Two numbers are often used to summarize a probability distribution for a random variable X. The *mean* is a measure of the center or middle of the probability distribution, and the *variance* is a measure of the dispersion, or variability in the distribution. These two measures do not uniquely identify a probability distribution. That is, two different distributions can have the same mean and variance. Still, these measures are simple, useful summaries of the probability distribution of X.

Mean, Variance, and Standard Deviation

The **mean** or **expected value** of the discrete random variable X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \sum_x xf(x) \tag{3.3}$$

The **variance** of X, denoted as σ^2 or $V(X)$, is

$$\sigma^2 = V(X) = E(X-\mu)^2 = \sum_x (x-\mu)^2 f(x) = \sum_x x^2 f(x) - \mu^2$$

The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$.

The mean of a discrete random variable X is a weighted average of the possible values of X with weights equal to the probabilities. If $f(x)$ is the probability mass function of a loading on a long, thin beam, $E(X)$ is the point at which the beam balances. Consequently, $E(X)$ describes the "center" of the distribution of X in a manner similar to the balance point of a loading. See Figure 3.4.

The variance of a random variable X is a measure of dispersion or scatter in the possible values for X. The standard deviation is also a widely used measure. The variance of X uses weight $f(x)$ as the multiplier of each possible squared deviation $(x-\mu)^2$. Figure 3.4 illustrates probability distributions with equal means but different variances. Properties of summations and the definition of μ can be used to show the equality of the formulas for variance in Equation 3.3.

$$\begin{aligned} V(X) &= \sum_x (x-\mu)^2 f(x) = \sum_x x^2 f(x) - 2\mu \sum_x xf(x) + \mu^2 \sum_x f(x) \\ &= \sum_x x^2 f(x) - 2\mu^2 + \mu^2 = \sum_x x^2 f(x) - \mu^2 \end{aligned}$$

Either formula for $V(x)$ can be used. Figure 3.5 illustrates that two probability distributions can differ even though they have identical means and variances.

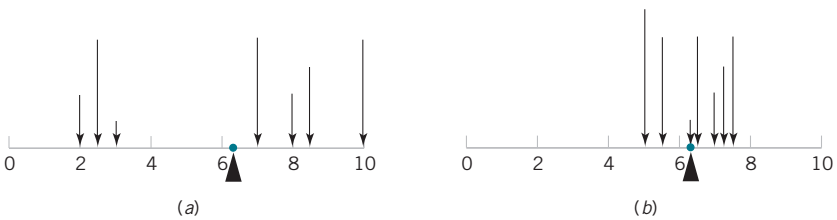


FIGURE 3.4

A probability distribution can be viewed as a loading with the mean equal to the balance point. Parts (a) and (b) illustrate equal means, but part (a) illustrates a larger variance.

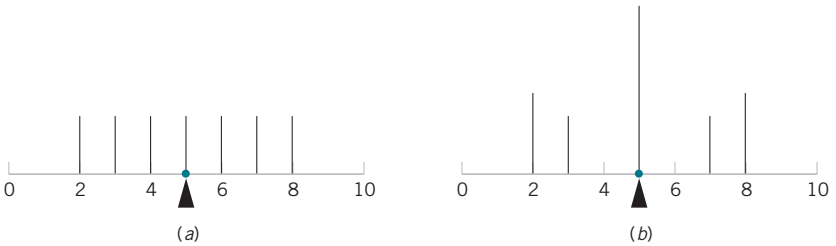


FIGURE 3.5

The probability distributions illustrated in parts (a) and (b) differ even though they have equal means and equal variances.

EXAMPLE 3.7 | Digital Channel

In Example 3.3, there is a chance that a bit transmitted through a digital transmission channel is received in error. Let X equal the number of bits in error in the next four bits transmitted. The possible values for X are $\{0, 1, 2, 3, 4\}$. Based on a model for the errors presented in the following section, probabilities for these values will be determined. Suppose that the probabilities are

$$P(X = 0) = 0.6561 \quad P(X = 2) = 0.0486 \quad P(X = 4) = 0.0001$$
$$P(X = 1) = 0.2916 \quad P(X = 3) = 0.0036$$

Now

$$\begin{aligned}\mu = E(X) &= 0f(0) + 1f(1) + 2f(2) + 3f(3) + 4f(4)\\ &= 0(0.6561) + 1(0.2916) + 2(0.0486) + 3(0.0036)\\ &\quad + 4(0.0001)\\ &= 0.4\end{aligned}$$

Although X never assumes the value 0.4, the weighted average of the possible values is 0.4.

To calculate $V(X)$, a table is convenient.

x	$x - 0.4$	$(x - 0.4)^2$	$f(x)$	$f(x)(x - 0.4)^2$
0	−0.4	0.16	0.6561	0.104976
1	0.6	0.36	0.2916	0.104976
2	1.6	2.56	0.0486	0.124416
3	2.6	6.76	0.0036	0.024336
4	3.6	12.96	0.0001	0.001296

$$V(X) = \sigma^2 = \sum_{i=1}^{5} f(x_i)(x_i - 0.4)^2 = 0.36$$

The alternative formula for variance could also be used to obtain the same result.

Practical Interpretation: The mean and variance summarize the distribution of a random variable. The mean is a weighted average of the values, and the variance measures the dispersion of the values from the mean. Different distributions may have the same mean and variance.

EXAMPLE 3.8 | Marketing

Two new product designs are to be compared on the basis of revenue potential. Marketing believes that the revenue from design A can be predicted quite accurately to be \$3 million. The revenue potential of design B is more difficult to assess. Marketing concludes that there is a probability of 0.3 that the revenue from design B will be \$7 million, but there is a 0.7 probability that the revenue will be only \$2 million. Which design do you prefer?

Let X denote the revenue from design A. Because there is no uncertainty in the revenue from design A, we can model the distribution of the random variable X as \$3 million with probability 1. Therefore, $E(X) = \$3$ million. Let Y denote the revenue from design B. The expected value of Y in millions of dollars is

$$E(Y) = \$7(0.3) + \$2(0.7) = \$3.5$$

Because $E(Y)$ exceeds $E(X)$, we might prefer design B. However, the variability of the result from design B is larger. That is,

$$\begin{aligned}\sigma^2 &= (7 - 3.5)^2(0.3) + (2 - 3.5)^2(0.7)\\ &= 5.25 \text{ millions of dollars squared}\end{aligned}$$

Because the units of the variables in this example are millions of dollars and because the variance of a random variable squares the deviations from the mean, the units of σ^2 are millions of dollars squared. These units make interpretation difficult.

Because the units of standard deviation are the same as the units of the random variable, the standard deviation σ is easier to interpret. Here $\sigma = \sqrt{5.25} = 2.29$ millions of dollars and σ is large relative to μ.

The variance of a random variable X can be considered to be the expected value of a specific function of X, namely, $h(X) = (X - \mu)^2$. In general, the expected value of any function $h(X)$ of a discrete random variable is defined in a similar manner.

Expected Value of a Function of a Discrete Random Variable

If X is a discrete random variable with probability mass function $f(x)$,

$$E[h(X)] = \sum_x h(x) f(x) \tag{3.4}$$

EXAMPLE 3.9 | Digital Channel

In Example 3.7, X is the number of bits in error in the next four bits transmitted. What is the expected value of the square of the number of bits in error? Now, $h(X) = X^2$. Therefore,

$$E[h(X)] = 0^2 \times 0.6561 + 1^2 \times 0.2916 + 2^2 \times 0.0486 + 3^2 \times 0.0036 + 4^2 \times 0.0001 = 0.52$$

Practical Interpretation: The expected value of a function of a random variable is simply a weighted average of the function evaluated at the values of the random variable.

In Example 3.9, the expected value of $h(X) = X^2$ does not equal $h[E(X)]$. However, in the special case that $h(X) = aX + b$ (for any constants a and b), the following can be shown from the properties of sums in Equation 3.4.

$$E(aX + b) = aE(X) + b$$

and

$$V(aX + b) = a^2 V(X)$$

In Example 3.8, suppose that the revenue for design B is increased by 10%. Let the random variable U denote the new revenue. Then $U = h(Y) = 1.1Y$ and

$$E(U) = 1.1E(Y) = 1.1(3.5) = 3.85$$

in millions of dollars and

$$V(U) = 1.1^2 V(Y) = 1.1^2(5.25) = 6.35$$

in millions of dollars squared.

3.4 Discrete Uniform Distribution

The simplest discrete random variable is one that assumes only a finite number of possible values, each with equal probability. A random variable X that assumes each of the values $x_1, x_2, \ldots, x_n$ with equal probability $1/n$ is frequently of interest.

Discrete Uniform Distribution

A random variable X has a **discrete uniform distribution** if each of the n values in its range, $x_1, x_2, \ldots, x_n$, has equal probability. Then

$$f(x_i) = \frac{1}{n} \tag{3.5}$$

EXAMPLE 3.10 | Serial Number

The first digit of a part's serial number is equally likely to be any one of the digits 0 through 9. If one part is selected randomly from a large batch and X is the first digit of the serial number, X has a discrete uniform distribution with probability 0.1 for each value in $R = \{0, 1, 2, \ldots, 9\}$. That is,

$$f(x) = 0.1$$

for each value in R. The probability mass function of X is shown in Figure 3.6.

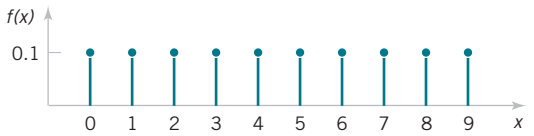


FIGURE 3.6
Probability mass function for a discrete uniform random variable.

Suppose that the range of the discrete random variable X equals the consecutive integers a, $a + 1$, $a + 2, \ldots, b$, for $a \leq b$. The range of X contains $b - a + 1$ values each with probability $1 / (b - a + 1)$. Now

$$\mu = \sum_{k=a}^{b} k \left(\frac{1}{b - a + 1} \right)$$

The algebraic identity $\sum_{k=a}^{b} k = \frac{b(b+1) - (a-1)a}{2}$ can be used to simplify the result to $\mu = (b + a) / 2$. The derivation of the variance is left as an exercise.

Mean and Variance

Suppose that X is a discrete uniform random variable on the consecutive integers a, $a + 1$, $a + 2, \ldots, b$, for $a \leq b$. The mean of X is

$$\mu = E(X) = \frac{b + a}{2}$$

The variance of X is

$$\sigma^2 = \frac{(b - a + 1)^2 - 1}{12} \tag{3.6}$$

EXAMPLE 3.11 | Number of Voice Lines

Let the random variable X denote the number of 48 voice lines that are used at a particular time. Assume that X is a discrete uniform random variable with a range of 0 to 48. Then,

$$E(X) = (48 + 0)/2 = 24$$

and

$$\sigma = \sqrt{\frac{(48 - 0 + 1)^2 - 1}{12}} = 14.14$$

Practical Interpretation: The average number of lines in use is 24, but the dispersion (as measured by σ) is large. Therefore, at many times far more or fewer than 24 lines are used.

Equation 3.6 is more useful than it might first appear. For example, suppose that the discrete uniform random variable Y has range $5, 10, \ldots, 30$. Then $Y = 5X$ where X has range $1, 2, \ldots, 6$.

The mean and variance of Y are obtained from the formulas for a linear function of X in Section 3.3 to be

$$E(Y) = 5E(X) = 5\left(\frac{1+6}{2}\right) = 17.5$$

$$V(Y) = 5^2 V(X) = 25\left[\frac{(6-1+1)^2-1}{12}\right] = 72.92$$

EXAMPLE 3.12 | Proportion of Voice Lines

Let the random variable Y denote the proportion of the 48 voice lines used at a particular time, and X denote the number of lines used at a particular time. Then $Y = X / 48$. Therefore,

$$E(Y) = E(X)/48 = 0.5$$

and

$$V(Y) = V(X)/48^2 = 0.087$$

3.5 Binomial Distribution

Consider the following random experiments and random variables:

1. Flip a coin 10 times. Let X = number of heads obtained.
2. A worn machine tool produces 1% defective parts. Let X = number of defective parts in the next 25 parts produced.
3. Each sample of air has a 10% chance of containing a particular rare molecule. Let X = the number of air samples that contain the rare molecule in the next 18 samples analyzed.
4. Of all bits transmitted through a digital transmission channel, 10% are received in error. Let X = the number of bits in error in the next five bits transmitted.
5. A multiple-choice test contains 10 questions, each with four choices, and you guess at each question. Let X = the number of questions answered correctly.
6. In the next 20 births at a hospital, let X = the number of female births.
7. Of all patients suffering a particular illness, 35% experience improvement from a medication. In the next 100 patients administered the medication, let X = the number of patients who experience improvement.

These examples illustrate that a general probability model that includes these experiments as particular cases would be very useful. Each of these random experiments can be thought of as consisting of a series of repeated, random trials: 10 flips of the coin in experiment 1, the production of 25 parts in experiment 2, and so forth. The random variable in each case is a count of the number of trials that meet a specified criterion. The outcome from each trial either meets the criterion that X counts or it does not; consequently, each trial can be summarized as resulting in either a success or a failure. For example, in the multiple-choice experiment, for each question, only the choice that is correct is considered a success. Choosing any one of the three incorrect choices results in the trial being summarized as a failure.

The terms *success* and *failure* are just labels. We can just as well use A and B or 0 or 1. Unfortunately, the usual labels can sometimes be misleading. In experiment 2, because X counts defective parts, the production of a defective part is called a *success*.

A trial with only two possible outcomes is used so frequently as a building block of a random experiment that it is called a **Bernoulli trial**. It is usually assumed that the trials that constitute the random experiment are *independent*. This implies that the outcome from one trial has no effect

on the outcome to be obtained from any other trial. Furthermore, it is often reasonable to assume that the *probability of a success in each trial is constant*. In the multiple-choice experiment, if the test taker has no knowledge of the material and just guesses at each question, we might assume that the probability of a correct answer is 1/4 for each question.

EXAMPLE 3.13 | Digital Channel

The chance that a bit transmitted through a digital transmission channel is received in error is 0.1. Also, assume that the transmission trials are independent. Let X = the number of bits in error in the next four bits transmitted. Determine $P(X = 2)$.

Let the letter E denote a bit in error, and let the letter O denote that the bit is okay, that is, received without error. We can represent the outcomes of this experiment as a list of four letters that indicate the bits that are in error and those that are okay. For example, the outcome $OEOE$ indicates that the second and fourth bits are in error and the other two bits are okay. The corresponding values for x are

Outcome	x	Outcome	x
$OOOO$	0	$EOOO$	1
$OOOE$	1	$EOOE$	2
$OOEO$	1	$EOEO$	2
$OOEE$	2	$EOEE$	3
$OEOO$	1	$EEOO$	2
$OEOE$	2	$EEOE$	3
$OEEO$	2	$EEEO$	3
$OEEE$	3	$EEEE$	4

The event that $X = 2$ consists of the six outcomes:

$$\{EEOO, EOEO, EOOE, OEEO, OEOE, OOEE\}$$

Using the assumption that the trials are independent, the probability of $\{EEOO\}$ is

$$P(EEOO) = P(E)P(E)P(O)P(O) = (0.1)^2(0.9)^2 = 0.0081$$

Also, any one of the six mutually exclusive outcomes for which $X = 2$ has the same probability of occurring. Therefore,

$$P(X = 2) = 6(0.0081) = 0.0486$$

In general, $P(X = x)$ = (number of outcomes that result in x errors) $\times (0.1)^x(0.9)^{4-x}$.

To complete a general probability formula, only an expression for the number of outcomes that contain x errors is needed. An outcome that contains x errors can be constructed by partitioning the four trials (letters) in the outcome into two groups. One group is of size x and contains the errors, and the other group is of size $n - x$ and consists of the trials that are okay. The number of ways of partitioning four trials into two groups, one of which is of size x, is $\binom{4}{x} = \frac{4!}{x!(4-x)!}$. Therefore, in this example,

$$P(X = x) = \binom{4}{x}(0.1)^x(0.9)^{4-x}$$

Notice that $\binom{4}{2} = 4!/[2!2!] = 6$, as found above. The probability mass function of X was shown in Example 3.3 and Figure 3.1.

The previous example motivates the following result.

Binomial Distribution

A random experiment consists of n Bernoulli trials such that

(1) The trials are independent.
(2) Each trial results in only two possible outcomes, labeled as "success" and "failure."
(3) The probability of a success in each trial, denoted as p, remains constant.

The random variable X that equals the number of trials that result in a success is a **binomial random variable** with parameters $0 < p < 1$ and $n = 1, 2, \ldots$. The probability mass function of X is

$$f(x) = \binom{n}{x} p^x(1-p)^{n-x} \qquad x = 0, 1, \ldots, n \tag{3.7}$$

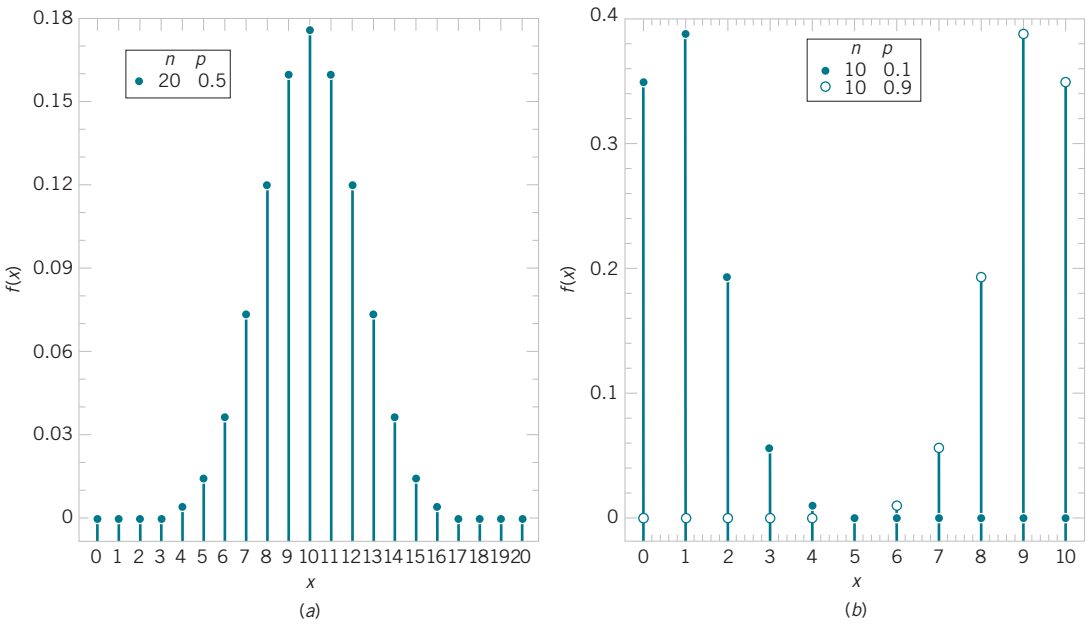


FIGURE 3.7

Binomial distributions for selected values of n and p.

As in Example 3.13, $\binom{n}{x}$ equals the total number of different sequences of trials that contain x successes and $n - x$ failures. The total number of different sequences that contain x successes and $n - x$ failures times the probability of each sequence equals $P(X = x)$.

The preceding probability expression is a very useful formula that can be applied in a number of examples. The name of the distribution is obtained from the *binomial expansion*. For constants a and b, the binomial expansion is

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Let p denote the probability of success on a single trial. Then by using the binomial expansion with $a = p$ and $b = 1 - p$, we see that the sum of the probabilities for a binomial random variable is 1. Furthermore, because each trial in the experiment is classified into two outcomes, {*success, failure*}, the distribution is called a "bi"-nomial. A more general distribution, for more than two outcomes, is the multinomial distribution.

Examples of binomial distributions are shown in Figure 3.7. For a fixed n, the distribution becomes more symmetric as p increases from 0 to 0.5 or decreases from 1 to 0.5. For a fixed p, the distribution becomes more symmetric as n increases.

EXAMPLE 3.14 | Binomial Coefficient

Several examples using the binomial coefficient $\binom{n}{x}$ follow.

$$\binom{10}{3} = 10!/[3!7!] = (10 \cdot 9 \cdot 8)/(3 \cdot 2) = 120$$

$$\binom{15}{10} = 15!/[10!5!] = (15 \cdot 14 \cdot 13 \cdot 12 \cdot 11)/(5 \cdot 4 \cdot 3 \cdot 2) = 3003$$

$$\binom{100}{4} = 100!/[4!96!] = (100 \cdot 99 \cdot 98 \cdot 97)/(4 \cdot 3 \cdot 2) = 3{,}921{,}225$$

Also recall that $0! = 1$.

EXAMPLE 3.15 | Organic Pollution

Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant. Find the probability that in the next 18 samples, exactly 2 contain the pollutant.

Let X = the number of samples that contain the pollutant in the next 18 samples analyzed. Then X is a binomial random variable with $p = 0.1$ and $n = 18$. Therefore,

$$P(X = 2) = \binom{18}{2}(0.1)^2(0.9)^{16}$$

Now $\binom{18}{2} = 18!/[2!16!] = 18(17)/2 = 153$. Therefore,

$$P(X = 2) = 153(0.1)^2(0.9)^{16} = 0.284$$

Determine the probability that at least four samples contain the pollutant. The requested probability is

$$P(X \geq 4) = \sum_{x=4}^{18} \binom{18}{x}(0.1)^x(0.9)^{18-x}$$

However, it is easier to use the complementary event,

$$P(X \geq 4) = 1 - P(X < 4) = 1 - \sum_{x=0}^{3} \binom{18}{x}(0.1)^x(0.9)^{18-x}$$
$$= 1 - [0.150 + 0.300 + 0.284 + 0.168] = 0.098$$

Determine the probability that $3 \leq X < 7$. Now

$$P(3 \leq X < 7) = \sum_{x=3}^{6} \binom{18}{x}(0.1)^x(0.9)^{18-x}$$
$$= 0.168 + 0.070 + 0.022 + 0.005$$
$$= 0.265$$

Practical Interpretation: Binomial random variables are used to model many physical systems and probabilities for all such models can be obtained from the binomial probability mass function.

A table of cumulative binomial probabilities is provided in Appendix A, and it can simplify some calculations. For example, the binomial distribution in Example 3.13 has $p = 0.1$ and $n = 4$. A probability such as $P(X = 2)$ can be calculated from the table as

$$P(X = 2) = P(X \leq 2) - P(X \leq 1) = 0.9963 - 0.9477 = 0.0486$$

and this agrees with the result obtained previously.

The mean and variance of a binomial random variable can be obtained from an analysis of the independent trials that comprise the binomial experiment. Define new random variables

$$X_i = \begin{cases} 1 & \text{if } i\text{th trial is a success} \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, \ldots, n$. Then

$$X = X_1 + X_2 + \cdots + X_n$$

Also, it is easy to derive the mean and variance of each X_i as

$$E(X_i) = 1p + 0(1 - p) = p$$

and

$$V(X_i) = (1 - p)^2 p + (0 - p)^2(1 - p) = p(1 - p)$$

Sums of random variables are discussed in Chapter 5, and there the intuitively reasonable result that

$$E(X) = E(X_1) + E(X_2) + \cdots + E(X_n)$$

is derived. Furthermore, for the independent trials of a binomial experiment, Chapter 5 also shows that

$$V(X) = V(X_1) + V(X_2) + \cdots + V(X_n)$$

Because $E(X_i) = p$ and $V(X_i) = p(1 - p)$, we obtain the solution $E(X) = np$ and $V(X) = np(1 - p)$.

Mean and Variance

If X is a binomial random variable with parameters p and n,

$$\mu = E(X) = np \quad \text{and} \quad \sigma^2 = V(X) = np(1-p) \tag{3.8}$$

EXAMPLE 3.16 | Mean and Variance

For the number of transmitted bits received in error in Example 3.13, $n = 4$ and $p = 0.1$, so

$$E(X) = 4(0.1) = 0.4 \quad \text{and} \quad V(X) = 4(0.1)(0.9) = 0.36$$

and these results match those obtained from a direct calculation in Example 3.7.

3.6 Geometric and Negative Binomial Distributions

Geometric Distribution Consider a random experiment that is closely related to the one used in the definition of a binomial distribution. Again, assume a series of Bernoulli trials (independent trials with constant probability p of a success on each trial). However, instead of a fixed number of trials, trials are conducted until a success is obtained. Let the random variable X denote the number of trials until the first success. Example 3.4 analyzed successive wafers until a large particle was detected. Then X is the number of wafers analyzed.

EXAMPLE 3.17 | Digital Channel

The probability that a bit transmitted through a digital transmission channel is received in error is 0.1. Assume that the transmissions are independent events, and let the random variable X denote the number of bits transmitted *until* the first error.

Then $P(X = 5)$ is the probability that the first four bits are transmitted correctly and the fifth bit is in error. This event can be denoted as $\{OOOOE\}$, where O denotes an okay bit. Because the trials are independent and the probability of a correct transmission is 0.9,

$$P(X = 5) = P(OOOOE) = 0.9^4 0.1 = 0.066$$

Note that there is some probability that X will equal any integer value. Also, if the first trial is a success, $X = 1$. Therefore, the range of X is $\{1, 2, 3, \ldots\}$, that is, all positive integers.

Geometric Distribution

In a series of Bernoulli trials (independent trials with constant probability p of a success), the random variable X that equals the number of trials until the first success is a **geometric random variable** with parameter $0 < p < 1$ and

$$f(x) = (1-p)^{x-1} p \qquad x = 1, 2, \ldots \tag{3.9}$$

Examples of the probability mass functions for geometric random variables are shown in Figure 3.8. Note that the height of the line at x is $(1 - p)$ times the height of the line at $x - 1$. That is, the probabilities decrease in a geometric progression. The distribution acquires its name from this result.

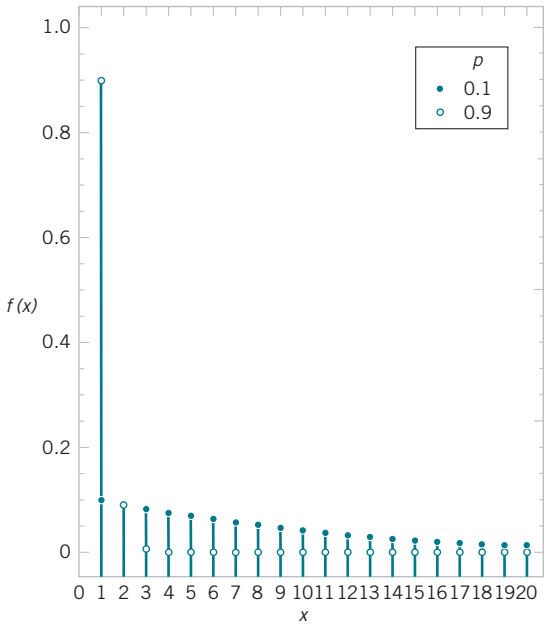


FIGURE 3.8
Geometric distributions for selected values of the parameter p.

EXAMPLE 3.18 | Wafer Contamination

The probability that a wafer contains a large particle of contamination is 0.01. If it is assumed that the wafers are independent, what is the probability that exactly 125 wafers need to be analyzed before a large particle is detected?

Let X denote the number of samples analyzed until a large particle is detected. Then X is a geometric random variable with $p = 0.01$. The requested probability is

$$P(X = 125) = (0.99)^{124}\, 0.01 = 0.0029$$

The mean of a geometric random variable is

$$\mu = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p\sum_{k=1}^{\infty} kq^{k-1}$$

where $q = p - 1$. The right-hand side of the previous equation is recognized to be the partial derivative with respect to q of

$$p\sum_{k=1}^{\infty} q^k = \frac{pq}{1-q}$$

where the last equality is obtained from the known sum of a geometric series. Therefore,

$$\mu = \frac{\partial}{\partial q}\left[\frac{pq}{1-q}\right] = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

and the mean is derived. To obtain the variance of a geometric random variable, we can first derive $E(X^2)$ by a similar approach. This can be obtained from partial second derivatives with respect to q. Then the formula $V(X) = E(X^2) - (E(X))^2$ is applied. The details are a bit more work, and this is left as a mind-expanding exercise.

Mean and Variance

If X is a geometric random variable with parameter p,

$$\mu = E(X) = 1/p \quad \text{and} \quad \sigma^2 = V(X) = (1-p)/p^2 \tag{3.10}$$

EXAMPLE 3.19 | Mean and Standard Deviation

Consider the transmission of bits in Example 3.17. Here $p = 0.1$. The mean number of transmissions until the first error is $1/0.1 = 10$. The standard deviation of the number of transmissions before the first error is

$$\sigma = [(1-0.1)/0.1^2]^{1/2} = 9.49$$

Practical Interpretation: The standard deviation here is approximately equal to the mean, and this occurs when p is small. Therefore, the actual number of trials until the first success may be much different from the mean when p is small.

Lack of Memory Property A geometric random variable has been defined as the number of trials until the first success. However, because the trials are independent, the count of the number of trials until the next success can be started at any trial without changing the probability distribution of the random variable. For example, if 100 bits are transmitted, the probability that the first error, after bit 100, occurs on bit 106 is the probability that the next six outcomes are *OOOOOE*. This probability is $(0.9)^5(0.1) = 0.059$, which is identical to the probability that the initial error occurs on bit 6.

The implication of using a geometric model is that the system presumably does not wear out. The probability of an error remains constant for all transmissions. In this sense, the geometric distribution is said to lack any memory. The **lack of memory property** is discussed again in the context of an exponential random variable in a later chapter.

EXAMPLE 3.20 | Lack of Memory Property

In Example 3.17, the probability that a bit is transmitted in error is equal to 0.1. Suppose that 50 bits have been transmitted. The mean number of bits until the next error is $1/0.1 = 10$—the same result as the mean number of bits until the first error.

Negative Binomial Distribution A generalization of a geometric distribution in which the random variable is the number of Bernoulli trials required to obtain r successes results in the **negative binomial distribution**.

EXAMPLE 3.21 | Digital Channel

As in Example 3.17, suppose that the probability that a bit transmitted through a digital transmission channel is received in error is 0.1. Assume that the transmissions are independent events, and let the random variable X denote the number of bits transmitted until the *fourth* error.

Then X has a negative binomial distribution with $r = 4$. Probabilities involving X can be found as follows. For example, $P(X = 10)$ is the probability that exactly three errors occur in the first 9 trials and then trial 10 results in the fourth error. The probability that exactly three errors occur in the first 9 trials is determined from the binomial distribution to be

$$\binom{9}{3}(0.1)^3(0.9)^6$$

Because the trials are independent, the probability that exactly three errors occur in the first 9 trials and trial 10 results in the fourth error is the product of the probabilities of these two events, namely,

$$\binom{9}{3}(0.1)^3(0.9)^6(0.1) = \binom{9}{3}(0.1)^4(0.9)^6$$

In general, probabilities for X can be determined as follows. Here $P(X = x)$ implies that $r - 1$ successes occur in the first $x - 1$ trials and the rth success occurs on trial x. The probability that $r - 1$ successes occur in the first $x - 1$ trials is obtained from the binomial distribution to be

$$\binom{x-1}{r-1} p^{r-1}(1-p)^{x-r}$$

for $r \leq x$. The probability that trial x is a success is p. Because the trials are independent, these probabilities are multiplied so that

$$P(X = x) = \binom{x-1}{r-1} p^{r-1}(1-p)^{x-r}p$$

This leads to the following result.

Negative Binomial Distribution

In a series of Bernoulli trials (independent trials with constant probability p of a success), the random variable X that equals the number of trials until r successes occur is a **negative binomial random variable** with parameters $0 < p < 1$ and $r = 1, 2, 3, \ldots$, and

$$f(x) = \binom{x-1}{r-1} (1-p)^{x-r}p^r \qquad x = r, r+1, r+2, \ldots \tag{3.11}$$

Because at least r trials are required to obtain r successes, the range of X is from r to ∞. In the special case that $r = 1$, a negative binomial random variable is a geometric random variable. Selected negative binomial distributions are illustrated in Figure 3.9.

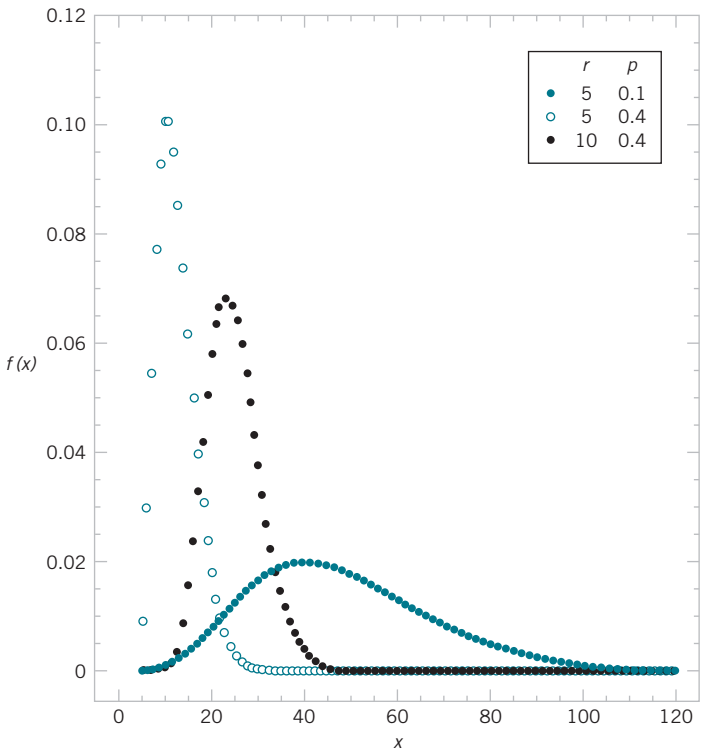


FIGURE 3.9

Negative binomial distributions for selected values of the parameters r and p.

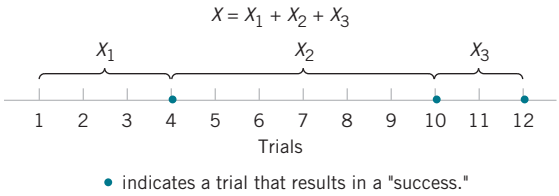


FIGURE 3.10
Negative binomial random variable represented as a sum of geometric random variables.

Let X denote the total number of trials required to obtain r successes. Let X_1 denote the number of trials required to obtain the first success, let X_2 denote the number of extra trials required to obtain the second success, let X_3 denote the number of extra trials to obtain the third success, and so forth. Then the total number of trials required to obtain r successes is $X = X_1 + X_2 + \cdots + X_r$. Because of the lack of memory property, each of the random variables $X_1, X_2, \ldots, X_r$ has a geometric distribution with the same value of p. Consequently, a negative binomial random variable can be interpreted as the sum of r geometric random variables. This concept is illustrated in Figure 3.10.

Recall that a binomial random variable is a count of the number of successes in n Bernoulli trials. That is, the number of trials is predetermined, and the number of successes is random. A negative binomial random variable is a count of the number of trials required to obtain r successes. That is, the number of successes is predetermined, and the number of trials is random. In this sense, a negative binomial random variable can be considered the opposite of a binomial random variable, although the name is based on the less intuitive result that the probability mass function can be written to appear similar to the binomial one if negative parameters are used.

The description of a negative binomial random variable as a sum of geometric random variables leads to the following results for the mean and variance.

Mean and Variance

If X is a negative binomial random variable with parameters p and r,

$$\mu = E(X) = r/p \quad \text{and} \quad \sigma^2 = V(X) = r(1-p)/p^2 \tag{3.12}$$

EXAMPLE 3.22 | Camera Flashes

Consider the time to recharge the flash in Example 3.1. The probability that a camera passes the test is 0.8, and the cameras perform independently. What is the probability that the third failure is obtained in five or fewer tests?

Let X denote the number of cameras tested until three failures have been obtained. The requested probability is $P(X \le 5)$. Here X has a negative binomial distribution with $p = 0.2$ and $r = 3$. Therefore,

$$P(X \le 5) = \sum_{x=3}^{5} \binom{x-1}{2} 0.2^3 (0.8)^{x-3}$$

$$= 0.2^3 + \binom{3}{2} 0.2^3(0.8) + \binom{4}{2} 0.2^3(0.8)^2 = 0.058$$

3.7 Hypergeometric Distribution

A day's production of 850 manufactured parts contains 50 parts that do not conform to customer requirements. Two parts are selected at random without replacement from the day's production. Let A and B denote the events that the first and second parts are nonconforming, respectively. From counting parts in the sample space, $P(B \mid A) = 49/849$ and $P(A) = 50/850$. Consequently, knowledge that the first part is nonconforming suggests that it is less likely that the second part selected is nonconforming.

Let X equal the number of nonconforming parts in the sample. Then,

$$P(X = 0) = P(\text{both parts conform}) = \frac{800}{850} \cdot \frac{799}{849} = 0.886$$

$P(X = 1) = P$ (first part selected conforms and the second part selected does not, or the first part selected does not and the second part selected conforms)

$$= \frac{800}{850} \cdot \frac{50}{849} + \frac{50}{850} \cdot \frac{800}{849} = 0.111$$

$$P(X = 2) = P(\text{both parts do not conform}) = \frac{50}{850} \cdot \frac{49}{849} = 0.003$$

This experiment is fundamentally different from the examples based on the binomial distribution. In this experiment, the trials are not independent. Note that, in the unusual case that each unit selected is replaced before the next selection, the trials are independent and there is a constant probability of a nonconforming part on each trial. Then the number of nonconforming parts in the sample is a binomial random variable.

But samples are often selected without replacement. Although probabilities can be determined by the preceding reasoning, a general formula for computing probabilities when samples are selected without replacement is quite useful. The counting rules presented in Chapter 2 can be used to justify the following formula.

Hypergeometric Distribution

A set of N objects contains

K objects classified as successes

$N - K$ objects classified as failures

A sample of size n objects is selected randomly (without replacement) from the N objects where $K \le N$ and $n \le N$.

The random variable X that equals the number of successes in the sample is a **hypergeometric random variable** and

$$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \qquad x = \max\{0, n + K - N\} \text{ to } \min\{K, n\} \tag{3.13}$$

The expression $\min\{K, n\}$ is used in the definition of the range of X because the maximum number of successes that can occur in the sample is the smaller of the sample size, n, and the number of successes available, K. Also, if $n + K > N$, at least $n + K - N$ successes must occur in the sample. Selected hypergeometric distributions are illustrated in Figure 3.11.

EXAMPLE 3.23 | Sampling Without Replacement

The computations at the start of this section can be reanalyzed by using the general expression in the definition of a hypergeometric random variable. That is,

$$P(X = 0) = \frac{\binom{50}{0}\binom{800}{2}}{\binom{850}{2}} = \frac{319{,}600}{360{,}825} = 0.886$$

$$P(X = 1) = \frac{\binom{50}{1}\binom{800}{1}}{\binom{850}{2}} = \frac{40{,}000}{360{,}825} = 0.111$$

$$P(X = 2) = \frac{\binom{50}{2}\binom{800}{0}}{\binom{850}{2}} = \frac{1{,}225}{360{,}825} = 0.003$$

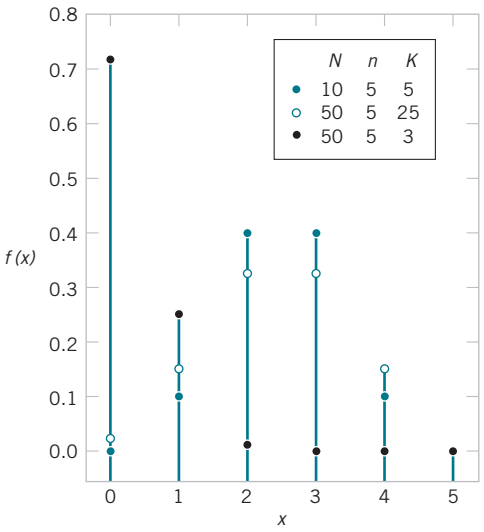


FIGURE 3.11
Hypergeometric distributions for selected values of parameters N, K, and n.

EXAMPLE 3.24 | Parts from Suppliers

A batch of parts contains 100 from a local supplier of circuit boards and 200 from a supplier in the next state. If four parts are selected randomly and without replacement, what is the probability they are all from the local supplier?

Let X equal the number of parts in the sample from the local supplier. Then X has a hypergeometric distribution and the requested probability is $P(X = 4)$. Consequently,

$$P(X = 4) = \frac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}} = 0.0119$$

What is the probability that two or more parts in the sample are from the local supplier?

$$P(X \geq 2) = \frac{\binom{100}{2}\binom{200}{2}}{\binom{300}{4}} + \frac{\binom{100}{3}\binom{200}{1}}{\binom{300}{4}} + \frac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}}$$
$$= 0.298 + 0.098 + 0.0119 = 0.407$$

What is the probability that at least one part in the sample is from the local supplier?

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{100}{0}\binom{200}{4}}{\binom{300}{4}} = 0.804$$

Practical Interpretation: Sampling without replacement is frequently used for inspection and the hypergeometric distribution simplifies the calculations.

The mean and variance of a hypergeometric random variable can be determined from the trials that compose the experiment. However, the trials are not independent, so the calculations are more difficult than for a binomial distribution. The results are stated as follows.

Mean and Variance

If X is a hypergeometric random variable with parameters N, K, and n, then

$$\mu = E(X) = np \qquad \text{and} \qquad \sigma^2 = V(X) = np(1-p)\left(\frac{N-n}{N-1}\right) \tag{3.14}$$

where $p = K/N$.

Here p is the proportion of successes in the set of N objects.

EXAMPLE 3.25 | Mean and Variance

In Example 3.24, the sample size is four. The random variable X is the number of parts in the sample from the local supplier. Then, $p = 100/300 = 1/3$. Therefore,

$$E(X) = 4\left(\frac{100}{300}\right) = 1.33$$

and

$$V(X) = 4\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)\left(\frac{300-4}{299}\right) = 0.88$$

For a hypergeometric random variable, $E(X)$ is similar to the mean of a binomial random variable. Also, $V(X)$ differs from the result for a binomial random variable only by the following term.

Finite Population Correction Factor

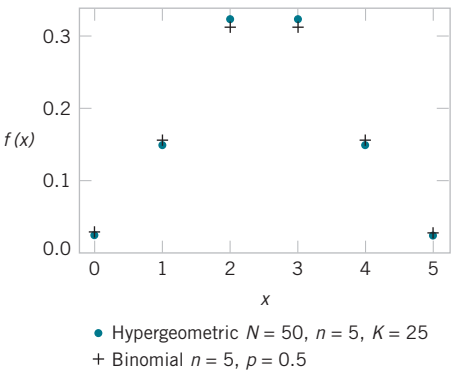
The term in the variance of a hypergeometric random variable

$$\frac{N-n}{N-1} \tag{3.15}$$

is called the **finite population correction factor**.

Sampling with replacement is equivalent to sampling from an infinite set because the proportion of success remains constant for every trial in the experiment. As mentioned previously, if sampling were done with replacement, X would be a binomial random variable and its variance would be $np(1-p)$. Consequently, the finite population correction represents the correction to the binomial variance that results because the sampling is without replacement from the finite set of size N.

If n is small relative to N, the correction is small and the hypergeometric distribution is similar to the binomial distribution. In this case, a binomial distribution can effectively approximate the hypergeometric distribution. A case is illustrated in Figure 3.12.



	0	1	2	3	4	5
Hypergeometric probability	0.025	0.149	0.326	0.326	0.149	0.025
Binomial probability	0.031	0.156	0.312	0.312	0.156	0.031

FIGURE 3.12

Comparison of hypergeometric and binomial distributions.

3.8 Poisson Distribution

A widely used distribution emerges from the concept that events occur randomly in an interval (or, more generally, in a region). The random variable of interest is the count of events that occur within the interval. Consider the following example.

EXAMPLE 3.26 | Wire Flaws

Flaws occur at random along the length of a thin copper wire. Let X denote the random variable that counts the number of flaws in a length of T millimeters of wire and suppose that the average number of flaws per millimeter is λ.

We expect $E(X) = \lambda T$ from the definition of λ. The probability distribution of X is determined as follows. Partition the length of wire into n subintervals of small length $\Delta t = T/n$ (say, one micrometer each). If the subintervals are chosen small enough, the probability that more than one flaw occurs in a subinterval is negligible. Furthermore, we can interpret the assumption that flaws occur at random to imply that every subinterval has the same probability of containing a flaw, say p. Also, the occurrence of a flaw in a subinterval is assumed to be independent of flaws in other subintervals.

Then we can model the distribution of X as approximately a binomial random variable. Each subinterval generates an event (flaw) or not. Therefore,

$$E(X) = \lambda T = np$$

and one can solve for p to obtain

$$p = \frac{\lambda T}{n}$$

From the approximate binomial distribution

$$P(X = x) \approx \binom{n}{x} p^x (1-p)^{n-x}$$

With small enough subintervals, n is large and p is small. Basic properties of limits can be used to show that as n increases

$$\binom{n}{x}\left(\frac{\lambda T}{n}\right)^x \to \frac{(\lambda T)^x}{x!}\left(1-\frac{\lambda T}{n}\right)^{-x} \to 1\left(1-\frac{\lambda T}{n}\right)^n \to e^{-\lambda T}$$

Therefore,

$$\lim_{n\to\infty} P(X = x) = \frac{e^{-\lambda T}(\lambda T)^x}{x!},\ x = 0, 1, 2, \ldots$$

Because the number of subintervals tends to infinity, the range of X (the number of flaws) can equal any nonnegative integer.

Example 3.26 can be generalized to include a broad array of random experiments. The interval that was partitioned was a length of wire. However, the same reasoning can be applied to an interval of time, an area, or a volume. For example, counts of (1) particles of contamination in semiconductor manufacturing, (2) flaws in rolls of textiles, (3) calls to a telephone exchange, (4) power outages, and (5) atomic particles emitted from a specimen have all been successfully modeled by the probability mass function in the following definition.

In general, consider subintervals of small length Δt and assume that as Δt tends to zero,

1. The probability of more than one event in a subinterval tends to zero.
2. The probability of one event in a subinterval tends to $\lambda \Delta t$.
3. The event in each subinterval is independent of other subintervals.

A random experiment with these properties is called a **Poisson process**.

These assumptions imply that the subintervals can be thought of as approximate independent Bernoulli trials with the number of trials equal to $n = T/\Delta t$ and success probability $p = \lambda \Delta t = \lambda T/n$. This leads to the following result.

Poisson Distribution

The random variable X that equals the number of events in a Poisson process is a **Poisson random variable** with parameter $0 < \lambda$, and

$$f(x) = \frac{e^{-\lambda T}(\lambda T)^x}{x!}\ x = 0, 1, 2, \ldots$$

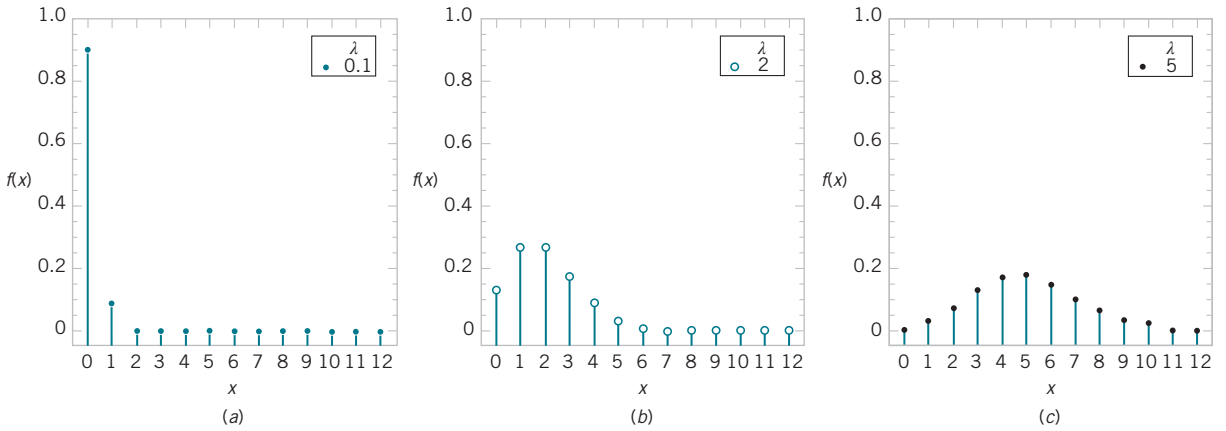


FIGURE 3.13

Poisson distributions for selected values of the parameters.

The sum of the probabilities is 1 because

$$\sum_{x=0}^{\infty} \frac{e^{-\lambda T}(\lambda T)^x}{x!} = e^{-\lambda T} \sum_{x=0}^{\infty} \frac{(\lambda T)^x}{x!}$$

and the summation on the right-hand side of the previous equation is recognized to be Taylor's expansion of e^x evaluated at λT. Therefore, the summation equals $e^{\lambda T}$ and the right-hand side equals 1. Figure 3.13 illustrates Poisson distributions for selected values of the parameter. The distribution is named after its developer Siméon-Denis Poisson.

Historically, the term *process* has been used to suggest the observation of a system over time. In our example with the copper wire, we showed that the Poisson distribution can also apply to lengths, and it also can be applied to areas.

The parameter λ is the mean number of events per unit length. It is important to *use consistent units* for λ and T. For example, if $\lambda = 2.3$ flaws per millimeter, then T should be expressed in millimeters. If $\lambda = 7.1$ flaws per square centimeter, then an area of 4.5 square inches should be expressed as $T = 4.5(2.54^2) = 29.03$ square centimeters.

EXAMPLE 3.27 | Calculations for Wire Flaws

For the case of the thin copper wire, suppose that the number of flaws follows a Poisson distribution with a mean of 2.3 flaws per millimeter.

Determine the probability of 10 flaws in 5 millimeters of wire. Let X denote the number of flaws in 5 millimeters of wire. Then, X has a Poisson distribution with

$$\lambda T = 2.3 \text{ flaws/mm} \times 5 \text{ mm} = 11.5 \text{ flaws}$$

Therefore,

$$P(X = 10) = e^{-11.5} \frac{11.5^{10}}{10!} = 0.113$$

Determine the probability of at least one flaw in 2 millimeters of wire. Let X denote the number of flaws in 2 millimeters of wire. Then X has a Poisson distribution with

$$\lambda T = 2.3 \text{ flaws/mm} \times 2 \text{ mm} = 4.6 \text{ flaws}$$

Therefore,

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-4.6} = 0.9899$$

Practical Interpretation: Given the assumptions for a Poisson process and a value for λ, probabilities can be calculated for intervals of arbitrary length. Such calculations are widely used to set product specifications, control processes, and plan resources.

The mean of a Poisson random variable is

$$E(X) = \sum_{x=1}^{\infty} x \frac{e^{-\lambda T}(\lambda T)^x}{x!} = \lambda T \sum_{x=1}^{\infty} \frac{e^{-\lambda T}(\lambda T)^{x-1}}{(x-1)!}$$

where the summation can start at $x = 1$ because the $x = 0$ term is zero. If a change of variable $y = x - 1$ is used, the summation on the right-hand side of the previous equation is recognized to be the sum of the probabilities of a Poisson random variable and this equals 1. Therefore, the previous equation simplifies to

$$E(X) = \lambda T$$

To obtain the variance of a Poisson random variable, we can start with $E(X^2)$ and this equals

$$E(X^2) = \sum_{x=1}^{\infty} x^2 \frac{e^{-\lambda T}(\lambda T)^x}{x!} = \lambda T \sum_{x=1}^{\infty} x \frac{e^{-\lambda T}(\lambda T)^{x-1}}{(x-1)!}$$

Write $x = (x - 1) + 1$ to obtain

$$E(X^2) = \lambda T \sum_{x=1}^{\infty} (x-1) \frac{e^{-\lambda T}(\lambda T)^{x-1}}{(x-1)!} + \lambda T \sum_{x=1}^{\infty} \frac{e^{-\lambda T}(\lambda T)^{x-1}}{(x-1)!}$$

The summation in the first term on the right-hand side of the previous equation is recognized to be the mean of X, which equals λT so that the first term is $(\lambda T)^2$. The summation in the second term on the right-hand side is recognized to be the sum of the probabilities, which equals 1. Therefore, the previous equation simplifies to $E(X^2) = (\lambda T)^2 + \lambda T$. Because the $V(X) = E(X^2) - (EX)^2$, we have

$$V(X) = (\lambda T)^2 + \lambda T - (\lambda T)^2 = \lambda T$$

and the variance is derived.

Mean and Variance

If X is a Poisson random variable over an interval of length T with parameter λ, then

$$\mu = E(X) = \lambda T \quad \text{and} \quad \sigma^2 = V(X) = \lambda T \tag{3.16}$$

The mean and variance of a Poisson random variable are equal. For example, if particle counts follow a Poisson distribution with a mean of 25 particles per square centimeter, the variance is also 25 and the standard deviation of the counts is 5 per square centimeter. Consequently, information on the variability is very easily obtained. Conversely, if the variance of count data is much greater than the mean of the same data, the Poisson distribution is not a good model for the distribution of the random variable.

Important Terms and Concepts

Bernoulli trial
Binomial distribution
Cumulative distribution function—discrete random variable
Discrete uniform distribution
Expected value of a function of a discrete random variable
Finite population correction factor
Geometric distribution
Hypergeometric distribution
Lack of memory property—discrete random variable
Mean—discrete random variable
Negative binomial distribution
Poisson distribution
Poisson process
Probability distribution—discrete random variable
Probability mass function
Standard deviation—discrete random variable
Variance—discrete random variable

CHAPTER 4

Continuous Random Variables and Probability Distributions

Wolfgang Amri/iStockphoto

LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Determine probabilities from probability density functions
2. Determine probabilities from cumulative distribution functions and cumulative distribution functions from probability density functions, and the reverse
3. Calculate means and variances for continuous random variables
4. Understand the assumptions for some common continuous probability distributions
5. Calculate probabilities, determine means and variances for some common continuous probability distributions
6. Approximate probabilities for binomial and Poisson distributions

CHAPTER OUTLINE

4.1 Probability Distributions and Probability Density Functions

4.2 Cumulative Distribution Functions

4.3 Mean and Variance of a Continuous Random Variable

4.4 Continuous Uniform Distribution

4.5 Normal Distribution

4.6 Normal Approximation to the Binomial and Poisson Distributions

4.7 Exponential Distribution

4.8 Erlang and Gamma Distributions

4.9 Weibull Distribution

4.10 Lognormal Distribution

4.11 Beta Distribution

The kinetic theory of gases provides a link between statistics and physical phenomena. Physicist James Maxwell used some basic assumptions to determine the distribution of molecular velocity in a gas at equilibrium. As a result of molecular collisions, all directions of rebound are equally likely. From this concept, he assumed equal probabilities for velocities in all the x, y, and z directions and independence of these components of velocity. This alone is sufficient to show that the probability distribution of the velocity in a particular direction x is the continuous probability distribution known as the normal distribution. This fundamental probability distribution can be derived from other approaches (such as the central limit theorem, to be discussed in a later chapter), but the kinetic theory may be the most parsimonious. This role for the normal distribution illustrates one example of the importance of continuous probability distributions within science and engineering.

4.1 Probability Distributions and Probability Density Functions

Suppose that a dimensional length is measured on a manufactured part selected from a day's production. In practice, there can be small variations in the measurements due to many causes, such as vibrations, temperature fluctuations, operator differences, calibrations, cutting tool wear, bearing wear, and raw material changes. In an experiment such as this, the measurement is naturally represented as a random variable X, and it is reasonable to model the range of possible values of X with an interval of real numbers. Recall from Chapter 2 that a **continuous random variable** is a random variable with an interval (either finite or infinite) of real numbers for its range. The model provides for any precision in length measurements.

Because the number of possible values of X is uncountably infinite, X has a distinctly different distribution from the discrete random variables studied previously. But as in the discrete case, many physical systems can be modeled by the same or similar continuous random variables. These random variables are described, and example computations of probabilities, means, and variances are provided in the sections of this chapter.

Density functions are commonly used in engineering to describe physical systems. For example, consider the density of a loading on a long, thin beam as shown in Figure 4.1. For any point x along the beam, the density can be described by a function (in grams/cm). Intervals with large loadings correspond to large values for the function. The total loading between points a and b is determined as the integral of the density function from a to b. This integral is the area under the density function over this interval, and it can be loosely interpreted as the sum of all the loadings over this interval.

Similarly, a **probability density function** $f(x)$ can be used to describe the probability distribution of a continuous random variable X. If an interval is likely to contain a value for X, its probability is large and it corresponds to large values for $f(x)$. The probability that X is between a and b is determined as the integral of $f(x)$ from a to b. See Figure 4.2.

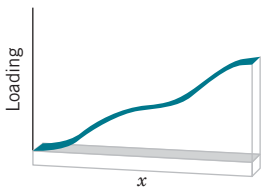


FIGURE 4.1
Density function of a loading on a long, thin beam.

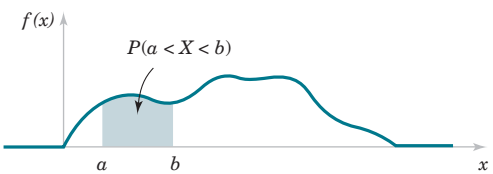


FIGURE 4.2
Probability determined from the area under $f(x)$.

Probability Density Function

For a continuous random variable X, a **probability density function** is a function such that

(1) $f(x) \geq 0$

(2) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

(3) $P(a \leq X \leq b) = \int_{a}^{b} f(x)\,dx =$ area under $f(x)$ from a to b for any a and b (4.1)

A probability density function provides a simple description of the probabilities associated with a random variable. As long as $f(x)$ is nonnegative and $\int_{-\infty}^{\infty} f(x)\,dx = 1$, $0 \leq P(a < X < b) \leq 1$ so that the probabilities are properly restricted. A probability density function is zero for x values that cannot occur, and it is assumed to be zero wherever it is not specifically defined.

A **histogram** is an approximation to a probability density function. See Figure 4.3. For each interval of the histogram, the area of the bar equals the relative frequency (proportion) of the measurements in the interval. The relative frequency is an estimate of the probability that a measurement falls in the interval. Similarly, the area under $f(x)$ over any interval equals the true probability that a measurement falls in the interval.

The important point is that $f(x)$ *is used to calculate an area* that represents the probability that X assumes a value in $[a, b]$. For the current measurement example, the probability that X results in [14 mA, 15 mA] is the integral of the probability density function of X over this interval. The probability that X results in [14.5 mA, 14.6 mA] is the integral of the same function, $f(x)$, over the smaller interval. By appropriate choice of the shape of $f(x)$, we can represent the probabilities associated with any continuous random variable X. The shape of $f(x)$ determines how the probability that X assumes a value in [14.5 mA, 14.6 mA] compares to the probability of any other interval of equal or different length.

For the density function of a loading on a long, thin beam, because every point has zero width, the loading at any point is zero. Similarly, for a continuous random variable X and *any* value x,

$$P(X = x) = 0$$

Based on this result, it might appear that our model of a continuous random variable is useless. However, in practice, when a particular current measurement, such as 14.47 milliamperes, is observed, this result can be interpreted as the rounded value of a current measurement that is actually in a range such as $14.465 \leq x \leq 14.475$. Therefore, the probability that the rounded value 14.47 is observed as the value for X is the probability that X assumes a value in the interval [14.465, 14.475], which is not zero. Similarly, because each point has zero probability, one need not distinguish between inequalities such as $<$ or $\leq$ for continuous random variables.

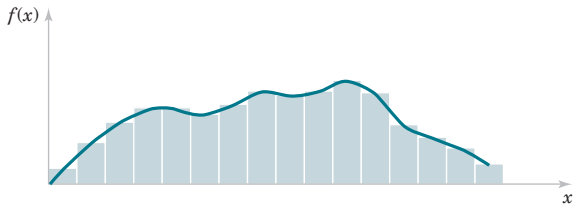


FIGURE 4.3
Histogram approximates a probability density function.

If X is a *continuous random variable*, for any x_1 and x_2,

$$P(x_1 \le X \le x_2) = P(x_1 < X \le x_2) = P(x_1 \le X < x_2) = P(x_1 < X < x_2) \qquad (4.2)$$

EXAMPLE 4.1 | Electric Current

Let the continuous random variable X denote the current measured in a thin copper wire in milliamperes. Assume that the range of X is [4.9, 5.1] mA, and assume that the probability density function of X is $f(x) = 5$ for $4.9 \le x \le 5.1$. What is the probability that a current measurement is less than 5 milliamperes?

The probability density function is shown in Figure 4.4. It is assumed that $f(x) = 0$ wherever it is not specifically defined. The shaded area in Figure 4.4 indicates the probability.

$$P(X < 5) = \int_{4.9}^{5} f(x)\,dx = \int_{4.9}^{5} 5\,dx = 0.5$$

As another example,

$$P(4.95 < X < 5.1) = \int_{4.95}^{5.1} f(x)\,dx = 0.75$$

EXAMPLE 4.2 | Hole Diameter

Let the continuous random variable X denote the diameter of a hole drilled in a sheet metal component. The target diameter is 12.5 millimeters. Most random disturbances to the process result in larger diameters. Historical data show that the distribution of X can be modeled by a probability density function $f(x) = 20e^{-20(x-12.5)}$, for $x \ge 12.5$.

If a part with a diameter greater than 12.60 mm is scrapped, what proportion of parts is scrapped? The density function and the requested probability are shown in Figure 4.5. A part is scrapped if $X > 12.60$. Now,

$$P(X > 12.60) = \int_{12.6}^{\infty} f(x)\,dx = \int_{12.6}^{\infty} 20e^{-20(x-12.5)}dx$$
$$= -e^{-20(x-12.5)}\Big|_{12.6}^{\infty} = 0.135$$

What proportion of parts is between 12.5 and 12.6 millimeters? Now,

$$P(12.5 < X < 12.6) = \int_{12.5}^{12.6} f(x)\,dx$$
$$= -e^{-20(x-12.5)}\Big|_{12.5}^{12.6} = 0.865$$

Because the total area under $f(x)$ equals 1, we can also calculate $P(12.5 < X < 12.6) = 1 - P(X > 12.6) = 1 - 0.135 = 0.865$.

Practical Interpretation: Because 0.135 is the proportion of parts with diameters greater than 12.60 mm, a large proportion of parts is scrapped. Process improvements are needed to increase the proportion of parts with dimensions near 12.50 mm.

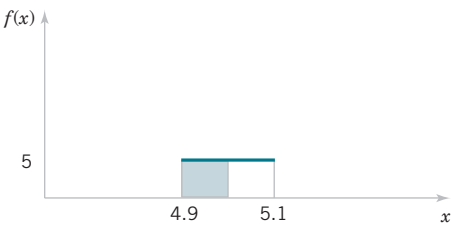


FIGURE 4.4
Probability density function for Example 4.1.

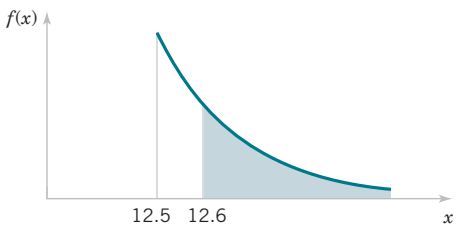


FIGURE 4.5
Probability density function for Example 4.2.

4.2 Cumulative Distribution Functions

An alternative method to describe the distribution of a discrete random variable can also be used for continuous random variables.

Cumulative Distribution Function

The **cumulative distribution function** of a continuous random variable X is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)\,du \tag{4.3}$$

for $-\infty < x < \infty$.

The cumulative distribution function is defined for all real numbers.

EXAMPLE 4.3 | Electric Current

For the copper current measurement in Example 4.1, the cumulative distribution function of the random variable X consists of three expressions. If $x < 4.9$, $f(x) = 0$. Therefore,

$$F(x) = 0, \quad \text{for} \quad x < 4.9$$

and

$$F(x) = \int_{4.9}^{x} f(u)\,du = 5x - 24.5, \quad \text{for} \quad 4.9 \le x < 5.1$$

Finally,

$$F(x) = \int_{4.9}^{x} f(u)\,du = 1, \quad \text{for} \quad 5.1 \le x$$

Therefore,

$$F(x) = \begin{cases} 0 & x < 4.9 \\ 5x - 24.5 & 4.9 \le x < 5.1 \\ 1 & 5.1 \ge x \end{cases}$$

The plot of $F(x)$ is shown in Figure 4.6.

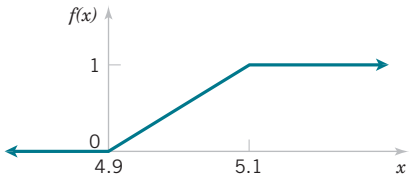


FIGURE 4.6

Cumulative distribution function for Example 4.3.

Notice that in the definition of $F(x)$, any $<$ can be changed to $\le$ and vice versa. That is, in Example 4.3 $F(x)$ can be defined as either $5x - 24.5$ or 0 at the end-point $x = 4.9$, and $F(x)$ can be defined as either $5x - 24.5$ or 1 at the end-point $x = 5.1$. In other words, $F(x)$ is a continuous function. For a discrete random variable, $F(x)$ is not a continuous function. Sometimes a continuous random variable is defined as one that has a continuous cumulative distribution function.

The probability density function of a continuous random variable can be determined from the cumulative distribution function by differentiating. The fundamental theorem of calculus states that

$$\frac{d}{dx}\int_{-\infty}^{x} f(u)du = f(x)$$

Probability Density Function from the Cumulative Distribution Function

Given $F(x)$,

$$f(x) = \frac{dF(x)}{dx}$$

as long as the derivative exists.

EXAMPLE 4.4 | Reaction Time

The time until a chemical reaction is complete (in milliseconds) is approximated by the cumulative distribution function

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-0.01x} & 0 \le x \end{cases}$$

Determine the probability density function of X. What proportion of reactions is complete within 200 milliseconds?

Using the result that the probability density function is the derivative of $F(x)$, we obtain

$$f(x) = \begin{cases} 0 & x < 0 \\ 0.01e^{-0.01x} & 0 \le x \end{cases}$$

The probability that a reaction completes within 200 milliseconds is

$$P(X < 200) = F(200) = 1 - e^{-2} = 0.8647$$

4.3 Mean and Variance of a Continuous Random Variable

The mean and variance can also be defined for a continuous random variable. Integration replaces summation in the discrete definitions. If a probability density function is viewed as a loading on a beam as in Figure 4.1, the mean is the balance point.

Mean and Variance

Suppose that X is a continuous random variable with probability density function $f(x)$. The **mean** or **expected value** of X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx \tag{4.4}$$

The **variance** of X, denoted as $V(X)$ or σ^2, is

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$.

The equivalence of the two formulas for variance can be derived from the same approach used for discrete random variables.

EXAMPLE 4.5 | Electric Current

For the copper current measurement in Example 4.1, the mean of X is

$$E(X) = \int_{4.9}^{5.1} xf(x)dx = \frac{5x^2}{2}\Big|_{4.9}^{5.1} = 5$$

The variance of X is

$$V(X) = \int_{4.9}^{5.1} (x-5)^2 f(x)dx = \frac{5(x-5)^3}{3}\Big|_{4.9}^{5.1} = 0.0033$$

The expected value of a function $h(X)$ of a continuous random variable is also defined in a straightforward manner.

Expected Value of a Function of a Continuous Random Variable

If X is a continuous random variable with probability density function $f(x)$,

$$E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx \tag{4.5}$$

In the special case that $h(X) = aX + b$ for any constants a and b, $E[h(X)] = aE(X) + b$. This can be shown from the properties of integrals.

EXAMPLE 4.6

In Example 4.1, X is the current measured in milliamperes. What is the expected value of power when the resistance is 100 ohms? Use the result that power in watts $P = 10^{-6}RI^2$, where I is the current in milliamperes and R is the resistance in ohms. Now, $h(X) = (10^{-6})100X^2$. Therefore,

$$E[h(X)] = 10^{-4}\int_{4.9}^{5.1} 5x^2dx = 0.0001\frac{x^3}{3}\Big|_{4.9}^{5.1} = 0.0025 \text{ watts}$$

4.4 Continuous Uniform Distribution

The simplest continuous distribution is analogous to its discrete counterpart.

Continuous Uniform Distribution

A continuous random variable X with probability density function

$$f(x) = 1/(b-a), \quad a \le x \le b \tag{4.6}$$

is a **continuous uniform random variable**.

The probability density function of a continuous uniform random variable is shown in Figure 4.7. The mean of the continuous uniform random variable X is

$$E(X) = \int_a^b \frac{x}{b-a}dx = \frac{0.5x^2}{b-a}\Big|_a^b = \frac{(a+b)}{2}$$

The variance of X is

$$V(X) = \int_a^b \frac{\left(x - \frac{(a+b)}{2}\right)^2}{b-a}dx = \frac{\left(x - \frac{(a+b)}{2}\right)^3}{3(b-a)}\Bigg|_a^b = \frac{(b-a)^2}{12}$$

These results are summarized as follows.

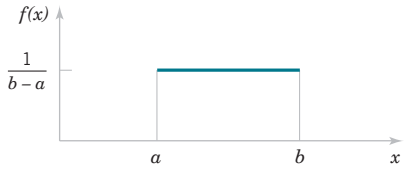


FIGURE 4.7
Continuous uniform probability density function.

Mean and Variance

If X is a continuous uniform random variable over $a \le x \le b$,

$$\mu = E(X) = \frac{a+b}{2} \quad \text{and} \quad \sigma^2 = V(X) = \frac{(b-a)^2}{12} \tag{4.7}$$

EXAMPLE 4.7 | Uniform Current

In Example 4.1, the random variable X has a continuous uniform distribution on [4.9, 5.1]. The probability density function of X is $f(x) = 5$, $4.9 \le x \le 5.1$.

What is the probability that a measurement of current is between 4.95 and 5.0 milliamperes? The requested probability is shown as the shaded area in Figure 4.8.

$$P(4.95 < x < 5.0) = \int_{4.95}^{5.0} f(x)dx = 5(0.05) = 0.25$$

The mean and variance formulas can be applied with $a = 4.9$ and $b = 5.1$. Therefore,

$$E(X) = 5 \text{ mA} \quad \text{and} \quad V(X) = 0.2^2/12 = 0.0033 \text{ mA}^2$$

Consequently, the standard deviation of X is 0.0577 mA.

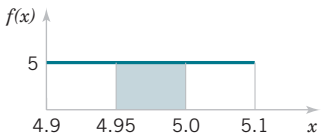


FIGURE 4.8
Probability for Example 4.7.

The cumulative distribution function of a continuous uniform random variable is obtained by integration. If $a < x < b$,

$$F(x) = \int_a^x \frac{1}{b-a} du = \frac{x-a}{b-a}$$

Therefore, the complete description of the cumulative distribution function of a continuous uniform random variable is

$$F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x < b \\ 1 & b \le x \end{cases}$$

An example of $F(x)$ for a continuous uniform random variable is shown in Figure 4.6.

4.5 Normal Distribution

Undoubtedly, the most widely used model for a continuous measurement is a **normal random variable**. Whenever a random experiment is replicated, the random variable that equals the average (or total) result over the replicates tends to have a normal distribution as the number of replicates becomes large. De Moivre presented this fundamental result, known as the **central limit theorem**, in 1733. Unfortunately, his work was lost for some time, and Gauss independently developed a normal distribution nearly 100 years later. Although De Moivre was later credited with the derivation, a normal distribution is also referred to as a **Gaussian distribution**.

When do we average (or total) results? Almost always. For example, an automotive engineer may plan a study to average pull-off force measurements from several connectors. If we assume that each measurement results from a replicate of a random experiment, the normal distribution can be used to make approximate conclusions about this average. These conclusions are the primary topics in the subsequent chapters of this book.

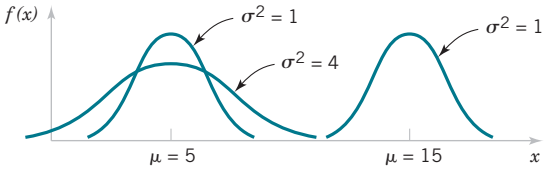


FIGURE 4.9

Normal probability density functions for selected values of the parameters μ and σ^2.

Furthermore, sometimes the central limit theorem is less obvious. For example, assume that the deviation (or error) in the length of a machined part is the sum of a large number of infinitesimal effects, such as temperature and humidity drifts, vibrations, cutting angle variations, cutting tool wear, bearing wear, rotational speed variations, mounting and fixture variations, variations in numerous raw material characteristics, and variations in levels of contamination. If the component errors are independent and equally likely to be positive or negative, the total error can be shown to have an approximate normal distribution. Furthermore, the normal distribution arises in the study of numerous basic physical phenomena. For example, physicist James Maxwell developed a normal distribution from simple assumptions regarding the velocities of molecules. Our objective now is to calculate probabilities for a normal random variable. The central limit theorem is stated more carefully in Chapter 5.

Figure 4.9 illustrates several normal probability density functions with selected values of the mean μ and the variance σ^2. Each has the characteristic symmetric bell-shaped curve, but the centers and dispersions differ. The following definition provides the formula for normal probability density functions.

Normal Distribution

A random variable X with probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty \tag{4.8}$$

is a **normal random variable** with parameters μ where $-\infty < \mu < \infty$ and $\sigma > 0$. Also,

$$E(X) = \mu \quad \text{and} \quad V(X) = \sigma^2 \tag{4.9}$$

and the notation $N(\mu, \sigma^2)$ is used to denote the distribution.

The mean and variance of X are shown to equal μ and σ^2, respectively, in an online exercise in Chapter 5.

EXAMPLE 4.8

Assume that the current measurements in a strip of wire follow a normal distribution with a mean of 10 milliamperes and a variance of 4 (milliamperes)^2. What is the probability that a measurement exceeds 13 milliamperes?

Let X denote the current in milliamperes. The requested probability can be represented as $P(X > 13)$. This probability is shown as the shaded area under the normal probability density function in Figure 4.10. Unfortunately, there is no closed-form expression for the integral of a normal probability density function, and probabilities based on the normal distribution are typically found numerically or from a table (that we introduce soon).

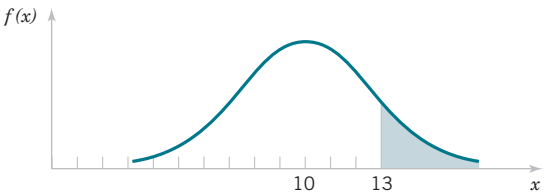


FIGURE 4.10

Probability that $X > 13$ for a normal random variable with $\mu = 10$ and $\sigma^2 = 4$.

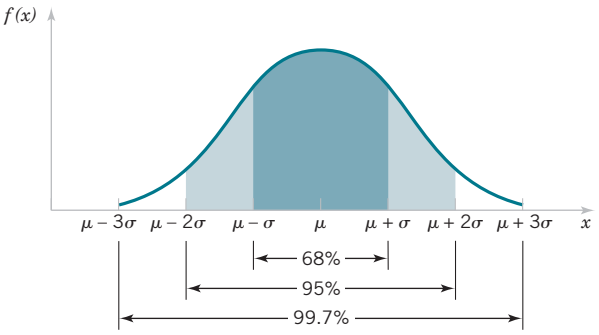


FIGURE 4.11
Probabilities associated with a normal distribution.

The following equations and Figure 4.11 summarize some useful results concerning a normal distribution. For any normal random variable,

$$P(\mu - \sigma < X < \mu + \sigma) = 0.6827$$
$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9545$$
$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$$

Also, from the symmetry of $P(X < \mu) = P(X > \mu) = 0.5$. Because $f(x)$ is positive for all x, this model assigns some probability to each interval of the real line. However, the probability density function decreases as x moves farther from μ. Consequently, the probability that a measurement falls far from μ is small, and at some distance from μ, the probability of an interval can be approximated as zero.

The area under a normal probability density function beyond 3σ from the mean is quite small. This fact is convenient for quick, rough sketches of a normal probability density function. The sketches help us determine probabilities. Because more than 0.9973 of the probability of a normal distribution is within the interval $(\mu - 3\sigma, \mu + 3\sigma)$, 6σ is often referred to as the **width** of a normal distribution. Advanced integration methods can be used to show that the area under the normal probability density function from $-\infty < x < \infty$ is 1.

Standard Normal Random Variable
A normal random variable with

$$\mu = 0 \quad \text{and} \quad \sigma^2 = 1$$

is called a **standard normal random variable** and is denoted as Z. The cumulative distribution function of a standard normal random variable is denoted as

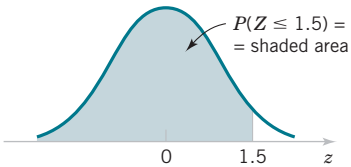
$$\Phi(z) = P(Z \leq z)$$

Appendix Table III provides cumulative probabilities for a standard normal random variable. Cumulative distribution functions for normal random variables are also widely available in computer packages. They can be used in the same manner as Appendix Table III to obtain probabilities for these random variables. The use of Table III is illustrated by the following example.

EXAMPLE 4.9 | Standard Normal Distribution

Assume that Z is a standard normal random variable. Appendix Table III provides probabilities of the form $\Phi(z) = P(Z \le z)$. The use of Table III to find $P(Z \le 1.5)$ is illustrated in Figure 4.12. Read down the z column to the row that equals 1.5. The probability is read from the adjacent column, labeled 0.00, to be 0.93319.

The column headings refer to the hundredths digit of the value of z in $P(Z \le z)$. For example, $P(Z \le 1.53)$ is found by reading down the z column to the row 1.5 and then selecting the probability from the column labeled 0.03 to be 0.93699.



z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50800	0.51197
⋮		⋮		
1.5	0.93319	0.93448	0.93574	0.93699

FIGURE 4.12

Standard normal probability density function.

Probabilities that are not of the form $P(Z \le z)$ are found by using the basic rules of probability and the symmetry of the normal distribution along with Appendix Table III. The following examples illustrate the method.

EXAMPLE 4.10 | Normal Distribution Calculations

The following calculations are shown pictorially in Figure 4.13. In practice, a probability is often rounded to one or two significant digits.

1. $P(Z > 1.26) = 1 - P(Z \le 1.26) = 1 - 0.89616 = 0.10384$.
2. $P(Z < -0.86) = 0.19490$.
3. $P(Z > -1.37) = P(Z < 1.37) = 0.91465$.
4. $P(-1.25 < Z < 0.37)$. This probability can be found from the difference of two areas, $P(Z < 0.37) - P(Z < -1.25)$. Now,

$$P(Z < 0.37) = 0.64431$$

and

$$P(Z < -1.25) = 0.10565$$

Therefore,

$$P(-1.25 < Z < 0.37) = 0.64431 - 0.10565 = 0.53866$$

5. $P(Z \le -4.6)$ cannot be found exactly from Appendix Table III. However, the last entry in the table can be used to find that $P(Z \le -3.99) = 0.00003$. Because $P(Z \le -4.6) < P(Z \le -3.99)$, $P(Z \le -4.6)$ is nearly zero.
6. Find the value z such that $P(Z > z) = 0.05$. This probability expression can be written as $P(Z \le z) = 0.95$. Now Table III is used in reverse. We search through the probabilities to find the value that corresponds to 0.95. The solution is illustrated in Figure 4.13. We do not find 0.95 exactly; the nearest value is 0.95053, corresponding to $z = 1.65$.
7. Find the value of z such that $P(-z < Z < z) = 0.99$. Because of the symmetry of the normal distribution, if the area of the shaded region in Figure 4.13(7) is to equal 0.99, the area in each tail of the distribution must equal 0.005. Therefore, the value for z corresponds to a probability of 0.995 in Table III. The nearest probability in Table III is 0.99506 when $z = 2.58$.

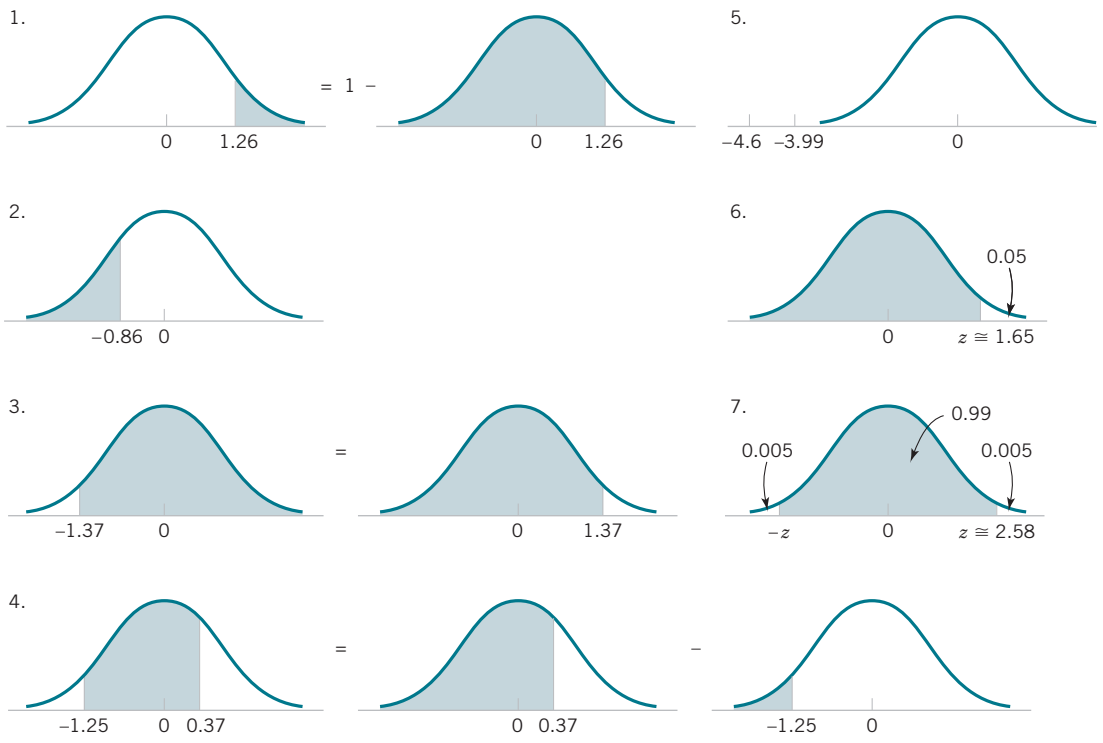


FIGURE 4.13

Graphical displays for standard normal distributions.

The cases in Example 4.10 show how to calculate probabilities for standard normal random variables. To use the same approach for an arbitrary normal random variable would require the availability of a separate table for every possible pair of values for μ and σ. Fortunately, all normal probability distributions are related algebraically, and Appendix Table III can be used to find the probabilities associated with an arbitrary normal random variable by first using a simple transformation.

Standardizing a Normal Random Variable

If X is a normal random variable with $E(X) = \mu$ and $V(X) = \sigma^2$, the random variable

$$Z = \frac{X - \mu}{\sigma} \tag{4.10}$$

is a normal random variable with $E(Z) = 0$ and $V(Z) = 1$. That is, Z is a standard normal random variable.

Creating a new random variable by this transformation is referred to as **standardizing**. The random variable Z represents the distance of X from its mean in terms of standard

deviations. It is the key step to calculating a probability for an arbitrary normal random variable.

EXAMPLE 4.11 | Normally Distributed Current

Suppose that the current measurements in a strip of wire are assumed to follow a normal distribution with a mean of 10 milliamperes and a variance of 4 (milliamperes)^2. What is the probability that a measurement exceeds 13 milliamperes?

Let X denote the current in milliamperes. The requested probability can be represented as $P(X > 13)$. Let $Z = (X - 10)/2$. The relationship between the several values of X and the transformed values of Z are shown in Figure 4.14. We note that $X > 13$ corresponds to $Z > 1.5$. Therefore, from Appendix Table III,

$$P(X > 13) = P(Z > 1.5) = 1 - P(Z \le 1.5)$$
$$= 1 - 0.93319 = 0.06681$$

Rather than using Figure 4.14, the probability can be found from the inequality $X > 13$. That is,

$$P(X > 13) = P\left(\frac{X - 10}{2} > \frac{13 - 10}{2}\right)$$
$$= P(Z > 1.5) = 0.06681$$

Practical Interpretation: Probabilities for any normal random variable can be computed with a simple transform to a standard normal random variable.

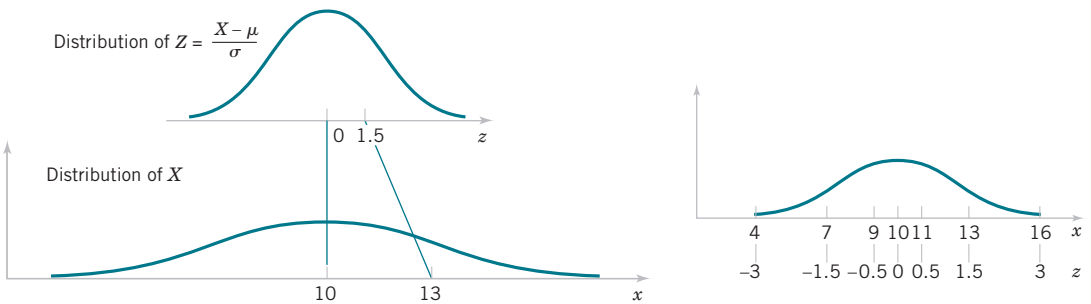


FIGURE 4.14

Standardizing a normal random variable.

In Example 4.11, the value 13 is transformed to 1.5 by standardizing, and 1.5 is often referred to as the ***z*-value** associated with a probability. The following summarizes the calculation of probabilities derived from normal random variables.

Standardizing to Calculate a Probability

Suppose that X is a normal random variable with mean μ and variance σ^2. Then,

$$P(X \le x) = P\left(\frac{X - \mu}{\sigma} \le \frac{x - \mu}{\sigma}\right) = P(Z \le z) \qquad (4.11)$$

where Z is a **standard normal random variable**, and $z = \dfrac{(x - \mu)}{\sigma}$ is the ***z*-value** obtained by **standardizing** X. The probability is obtained by using Appendix Table III with $z = (x - \mu)/\sigma$.

EXAMPLE 4.12 | Normally Distributed Current

Continuing Example 4.11, what is the probability that a current measurement is between 9 and 11 milliamperes? From Figure 4.14, or by proceeding algebraically, we have

$$\begin{aligned} P(9 < X < 11) &= P\left(\frac{9-10}{2} < \frac{X-10}{2} < \frac{11-10}{2}\right) \\ &= P(-0.5 < Z < 0.5) \\ &= P(Z < 0.5) - P(Z < -0.5) \\ &= 0.69146 - 0.30854 = 0.38292 \end{aligned}$$

Determine the value for which the probability that a current measurement is less than this value is 0.98. The requested value is shown graphically in Figure 4.15. We need the value of x such that $P(X < x) = 0.98$. By standardizing, this probability expression can be written as

$$\begin{aligned} P(X < x) &= P\left(\frac{X-10}{2} < \frac{x-10}{2}\right) \\ &= P\left(Z < \frac{x-10}{2}\right) = 0.98 \end{aligned}$$

Appendix Table III is used to find the z-value such that $P(Z < z) = 0.98$. The nearest probability from Table III results in

$$P(Z < 2.06) = 0.980301$$

Therefore, $(x - 10)/2 = 2.06$, and the standardizing transformation is used in reverse to solve for x. The result is

$$x = 2(2.06) + 10 = 14.1 \text{ mA}$$

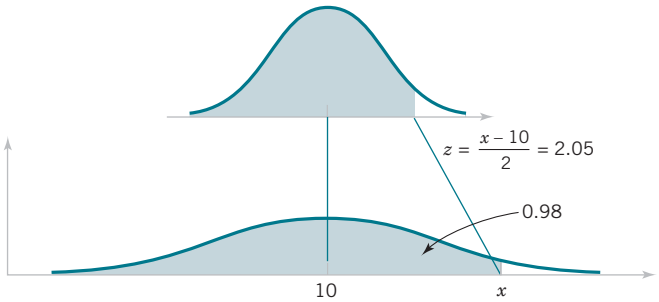


FIGURE 4.15

Determining the value of x to meet a specified probability.

4.6 Normal Approximation to the Binomial and Poisson Distributions

We began our section on the normal distribution with the central limit theorem and the normal distribution as an approximation to a random variable with a large number of trials. Consequently, it should not be surprising to learn that the normal distribution can be used to approximate binomial probabilities for cases in which n is large. The following example illustrates that for many physical systems, the binomial model is appropriate with an extremely large value for n. In these cases, it is difficult to calculate probabilities by using the binomial distribution. Fortunately, the **normal approximation** is most effective in these cases. An illustration is provided in Figure 4.16. The area of each bar equals the binomial probability of x. Notice that the area of bars can be approximated by areas under the normal probability density function.

From Figure 4.16, it can be seen that a probability such as $P(3 \le X \le 7)$ is better approximated by the area under the normal curve from 2.5 to 7.5. Consequently, a modified interval is used to better compensate for the difference between the continuous normal distribution and the discrete binomial distribution. This modification is called a **continuity correction**.

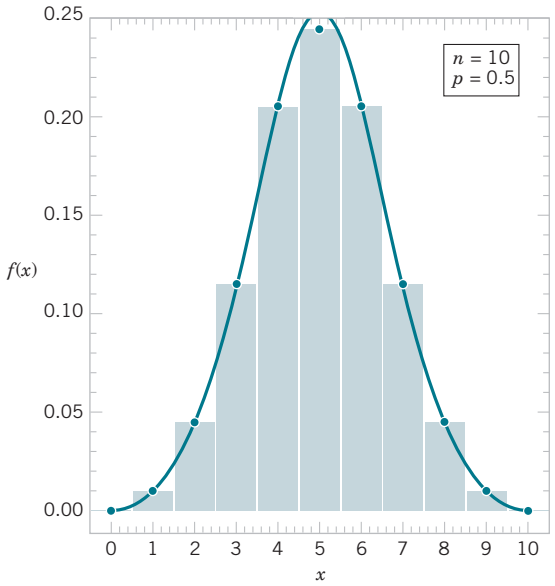


FIGURE 4.16
Normal approximation to the binomial distribution.

EXAMPLE 4.13

Assume that in a digital communication channel, the number of bits received in error can be modeled by a binomial random variable, and assume that the probability that a bit is received in error is 1×10^{-5}. If 16 million bits are transmitted, what is the probability that 150 or fewer errors occur?

Let the random variable X denote the number of errors. Then X is a binomial random variable and

$$P(X \le 150) = \sum_{x=0}^{150} \binom{16{,}000{,}000}{x} (10^{-5})^x (1 - 10^{-5})^{16{,}000{,}000 - x}$$

Practical Interpretation: Clearly, this probability is difficult to compute. Fortunately, the normal distribution can be used to provide an excellent approximation in this example.

Normal Approximation to the Binomial Distribution

If X is a binomial random variable with parameters n and p,

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \tag{4.12}$$

is approximately a standard normal random variable. To approximate a binomial probability with a normal distribution, a **continuity correction** is applied as follows:

$$P(X \le x) = P(X \le x + 0.5) \approx P\left(Z \le \frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

and

$$P(x \le X) = P(x - 0.5 \le X) \approx P\left(\frac{x - 0.5 - np}{\sqrt{np(1-p)}} \le Z\right)$$

The approximation is good for $np > 5$ and $n(1-p) > 5$.

Recall that for a binomial variable X, $E(X) = np$ and $V(X) = np(1-p)$. Consequently, the expression in Equation 4.12 is nothing more than the formula for standardizing the random variable X. Probabilities involving X can be approximated by using a standard normal distribution. The approximation is good when n is large relative to p.

A way to remember the continuity correction is to write the probability in terms of $\leq$ or $\geq$ and then add or subtract the 0.5 correction factor to make the probability greater.

EXAMPLE 4.14

The digital communication problem in Example 4.13 is solved as follows:

$$P(X \leq 150) = P(X \leq 150.5)$$
$$= P\left(\frac{X - 160}{\sqrt{160(1-10^{-5})}} \leq \frac{150.5 - 160}{\sqrt{160(1-10^{-5})}}\right)$$
$$\approx P(Z \leq -0.75) = 0.227$$

Because $np = (16 \times 10^6)(1 \times 10^{-5}) = 160$ and $n(1-p)$ is much larger, the approximation is expected to work well in this case.

Practical Interpretation: Binomial probabilities that are difficult to compute exactly can be approximated with easy-to-compute probabilities based on the normal distribution.

EXAMPLE 4.15 | Normal Approximation to Binomial

Again consider the transmission of bits in Example 4.14. To judge how well the normal approximation works, assume that only $n = 50$ bits are to be transmitted and that the probability of an error is $p = 0.1$. The exact probability that two or fewer errors occur is

$$P(X \leq 2) = \binom{50}{0} 0.9^{50} + \binom{50}{1} 0.1(0.9^{49}) + \binom{50}{2} 0.1^2(0.9^{48}) = 0.112$$

Based on the normal approximation,

$$P(X \leq 2) = P\left(\frac{X - 5}{\sqrt{50(0.1)(0.9)}} \leq \frac{2.5 - 5}{\sqrt{50(0.1)(0.9)}}\right)$$
$$\approx P(Z < -1.18) = 0.119$$

We can also approximate $P(X = 5)$ as

$$P(5 \leq X \leq 5) = P(4.5 \leq X \leq 5.5)$$
$$\approx P\left(\frac{4.5 - 5}{2.12} \leq Z \leq \frac{5.5 - 5}{2.12}\right)$$
$$= P(-0.24 \leq Z \leq 0.24) = 0.19$$

and this compares well with the exact answer of 0.1849.

Practical Interpretation: Even for a sample as small as 50 bits, the normal approximation is reasonable, when $p = 0.1$.

The correction factor is used to improve the approximation. However, if np or $n(1-p)$ is small, the binomial distribution is quite skewed and the symmetric normal distribution is not a good approximation. Two cases are illustrated in Figure 4.17.

Recall that the binomial distribution is a satisfactory approximation to the hypergeometric distribution when n, the sample size, is small relative to N, the size of the population from which the sample is selected. A rule of thumb is that the binomial approximation is effective if $n/N < 0.1$. Recall that for a hypergeometric distribution, p is defined as K/N. That is, p is interpreted as the number of successes in the population. Therefore, the normal distribution can provide an effective approximation of hypergeometric probabilities when $n/N < 0.1$, $np > 5$, and $n(1-p) > 5$. Figure 4.18 provides a summary of these guidelines.

Recall that the Poisson distribution was developed as the limit of a binomial distribution as the number of trials increased to infinity. Consequently, it should not be surprising to find that the normal distribution can also be used to approximate probabilities of a Poisson random variable.

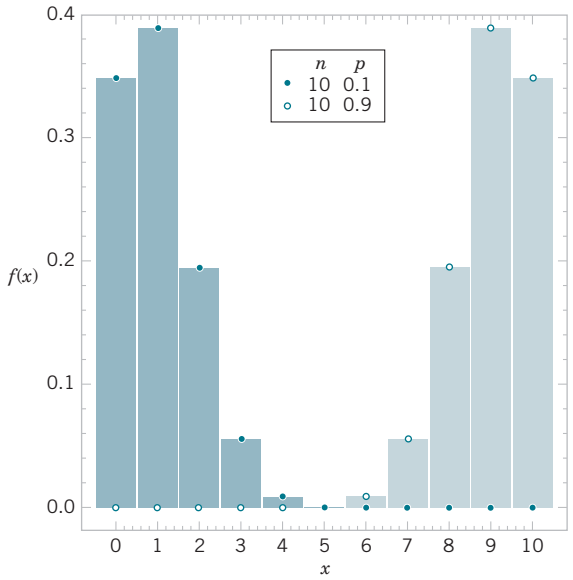


FIGURE 4.17

Binomial distribution is not symmetrical if p is near 0 or 1.

Normal Approximation to the Poisson Distribution

If X is a Poisson random variable with $E(X) = \lambda$ and $V(X) = \lambda$,

$$Z = \frac{X - \lambda}{\sqrt{\lambda}} \tag{4.13}$$

is approximately a standard normal random variable. The same continuity correction used for the binomial distribution can also be applied. The approximation is good for

$$\lambda > 5$$

EXAMPLE 4.16 | Normal Approximation to Poisson

Assume that the number of asbestos particles in a squared meter of dust on a surface follows a Poisson distribution with a mean of 1000. If a squared meter of dust is analyzed, what is the probability that 950 or fewer particles are found?

This probability can be expressed exactly as

$$P(X \le 950) = \sum_{x=0}^{950} \frac{e^{-1000}1000^x}{x!}$$

The computational difficulty is clear. The probability can be approximated as

$$P(X \le 950) = P(X \le 950.5) \approx P\left(Z \le \frac{950.5 - 1000}{\sqrt{1000}}\right)$$
$$= P(Z \le -1.57) = 0.058$$

Practical Interpretation: Poisson probabilities that are difficult to compute exactly can be approximated with easy-to-compute probabilities based on the normal distribution.

Hypergeometric distribution $\overset{\approx}{\underset{\frac{n}{N} < 0.1}{}}$ Binomial distribution $\overset{\approx}{\underset{\substack{np > 5 \\ n(1-p) > 5}}{}}$ Normal distribution

FIGURE 4.18

Conditions for approximating hypergeometric and binomial probabilities.

4.7 Exponential Distribution

The discussion of the Poisson distribution defined a random variable to be the number of flaws along a length of copper wire. The distance between flaws is another random variable that is often of interest. Let the random variable X denote the length from any starting point on the wire until a flaw is detected. As you might expect, the distribution of X can be obtained from knowledge of the distribution of the number of flaws. The key to the relationship is the following concept. The distance to the first flaw exceeds 3 millimeters if and only if there are no flaws within a length of 3 millimeters—simple but sufficient for an analysis of the distribution of X.

In general, let the random variable N denote the number of flaws in x millimeters of wire. If the mean number of flaws is λ per millimeter, N has a Poisson distribution with mean λx. We assume that the wire is longer than the value of x. Now,

$$P(X > x) = P(N = 0) = \frac{e^{-\lambda x}(\lambda x)^0}{0!} = e^{-\lambda x}$$

Therefore,

$$F(x) = P(X \leq x) = 1 - e^{-\lambda x}, \quad x \geq 0$$

is the cumulative distribution function of X. By differentiating $F(x)$, the probability density function of X is calculated to be

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

The derivation of the distribution of X depends only on the assumption that the flaws in the wire follow a **Poisson process**. Also, the starting point for measuring X does not matter because the probability of the number of flaws in an interval of a Poisson process depends only on the length of the interval, not on the location. For any Poisson process, the following general result applies.

Exponential Distribution

The random variable X that equals the distance between successive events from a Poisson process with mean number of events $\lambda > 0$ per unit interval is an **exponential random variable** with parameter λ. The probability density function of X is

$$f(x) = \lambda e^{-\lambda x} \quad \text{for} \quad 0 \leq x < \infty \tag{4.14}$$

The exponential distribution obtains its name from the exponential function in the probability density function. See plots of the exponential distribution for selected values of λ in Figure 4.19. For any value of λ, the exponential distribution is quite skewed. The following results are easily obtained and are left as an exercise.

Mean and Variance

If the random variable X has an exponential distribution with parameter λ,

$$\mu = E(X) = \frac{1}{\lambda} \quad \text{and} \quad \sigma^2 = V(X) = \frac{1}{\lambda^2} \tag{4.15}$$

It is important to *use consistent units* to express intervals, X, and λ. The following example illustrates unit conversions.

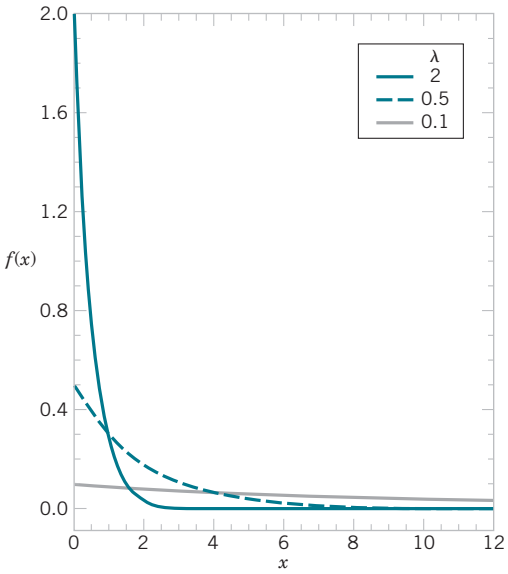


FIGURE 4.19
Exponential probability density functions for selected values of λ.

EXAMPLE 4.17 | Computer Usage

In a large corporate computer network, user log-ons to the system can be modeled as a Poisson process with a mean of 25 log-ons per hour. What is the probability that there are no log-ons in an interval of 6 minutes?

Let X denote the time in hours from the start of the interval until the first log-on. Then X has an exponential distribution with $\lambda = 25$ log-ons per hour. We are interested in the probability that X exceeds 6 minutes. Because λ is given in log-ons per hour, we express all time units in hours. That is, 6 minutes = 0.1 hour. Therefore,

$$P(X > 0.1) = \int_{0.1}^{\infty} 25e^{-25x}dx = e^{-25(0.1)} = 0.082$$

The cumulative distribution function can also be used to obtain the same result as follows:

$$P(X > 0.1) = 1 - F(0.1) = e^{-25(0.1)}$$

An identical answer is obtained by expressing the mean number of log-ons as 0.417 log-ons per minute and computing the probability that the time until the next log-on exceeds 6 minutes. Try it.

What is the probability that the time until the next log-on is between 2 and 3 minutes? Upon converting all units to hours,

$$P(0.033 < X < 0.05) = \int_{0.033}^{0.05} 25e^{-25x}dx$$
$$= -e^{-25x}\Big|_{0.033}^{0.05} = 0.152$$

An alternative solution is

$$P(0.033 < X < 0.05) = F(0.05) - F(0.033) = 0.152$$

Determine the interval of time such that the probability that no log-on occurs in the interval is 0.90. The question asks for the length of time x such that $P(X > x) = 0.90$. Now,

$$P(X > x) = e^{-25x} = 0.90$$

Take the (natural) log of both sides to obtain $-25x = \ln(0.90) = -0.1054$. Therefore,

$$x = 0.00421 \text{ hour} = 0.25 \text{ minute}$$

Furthermore, the mean time until the next log-on is

$$\mu = 1/25 = 0.04 \text{ hour} = 2.4 \text{ minutes}$$

The standard deviation of the time until the next log-on is

$$\sigma = 1/25 \text{ hours} = 2.4 \text{ minutes}$$

Practical Interpretation: Organizations make wide use of probabilities for exponential random variables to evaluate resources and staffing levels to meet customer service needs.

In Example 4.17, the probability that there are no log-ons in a six-minute interval is 0.082 regardless of the starting time of the interval. A Poisson process assumes that events occur uniformly throughout the interval of observation; that is, there is no clustering of events. If the log-ons are well modeled by a Poisson process, the probability that the first log-on after noon occurs after 12:06 P.M. is the same as the probability that the first log-on after 3:00 P.M. occurs after 3:06 P.M. And if someone logs on at 2:22 P.M., the probability that the next log-on occurs after 2:28 P.M. is still 0.082.

Our starting point for observing the system does not matter. However, if high-use periods occur during the day, such as right after 8:00 A.M., followed by a period of low use, a Poisson process is not an appropriate model for log-ons and the distribution is not appropriate for computing probabilities. It might be reasonable to model each of the high- and low-use periods by a separate Poisson process, employing a larger value for λ during the high-use periods and a smaller value otherwise. Then an exponential distribution with the corresponding value of λ can be used to calculate log-on probabilities for the high- and low-use periods.

Lack of Memory Property An even more interesting property of an exponential random variable concerns conditional probabilities.

EXAMPLE 4.18 | Lack of Memory Property

Let X denote the time between detections of a particle with a Geiger counter and assume that X has an exponential distribution with $E(X) = 1.4$ minutes. The probability that we detect a particle within 0.5 minute of starting the counter is

$$P(X < 0.5) = F(0.5) = 1 - e^{-0.5/1.4} = 0.30$$

Now, suppose that we turn on the Geiger counter and wait 3 minutes without detecting a particle. What is the probability that a particle is detected in the next 30 seconds?

Because we have already been waiting for 3 minutes, we feel that a detection is "due." That is, the probability of a detection in the next 30 seconds should be higher than 0.3. However, for an exponential distribution, this is not true. The requested probability can be expressed as the conditional probability that $P(X < 3.5 \mid X > 3)$. From the definition of conditional probability,

$$P(X < 3.5 \mid X > 3) = P(3 < X < 3.5)/P(X > 3)$$

where

$$P(3 < X < 3.5) = F(3.5) - F(3)$$
$$= [1 - e^{-3.5/1.4}] - [1 - e^{-3/1.4}] = 0.035$$

and

$$P(X > 3) = 1 - F(3) = e^{-3/1.4} = 0.117$$

Therefore,

$$P(X < 3.5 \mid X > 3) = 0.035/0.117 = 0.30$$

Practical Interpretation: After waiting for 3 minutes without a detection, the probability of a detection in the next 30 seconds is the same as the probability of a detection in the 30 seconds immediately after starting the counter. The fact that we have waited 3 minutes without a detection does not change the probability of a detection in the next 30 seconds.

Example 4.18 illustrates the **lack of memory property** of an exponential random variable, and a general statement of the property follows. In fact, the exponential distribution is the only continuous distribution with this property.

Lack of Memory Property

For an exponential random variable X,

$$P(X < t_1 + t_2 \mid X > t_1) = P(X < t_2) \tag{4.16}$$

Figure 4.20 graphically illustrates the lack of memory property. The area of region A divided by the total area under the probability density function ($A + B + C + D = 1$) equals $P(X < t_2)$.

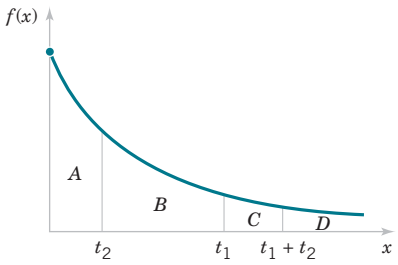


FIGURE 4.20
Lack of memory property of an exponential distribution.

The area of region C divided by the area $C + D$ equals $P(X < t_1 + t_2 \mid X > t_1)$. The lack of memory property implies that the proportion of the total area that is in A equals the proportion of the area in C and D that is in C.

The lack of memory property is not so surprising when we consider the development of a Poisson process. In that development, we assumed that an interval could be partitioned into small intervals that were independent. These subintervals are similar to independent Bernoulli trials that comprise a binomial experiment; knowledge of previous results does not affect the probabilities of events in future subintervals. An exponential random variable is the continuous analog of a geometric random variable, and it shares a similar lack of memory property.

The exponential distribution is often used in reliability studies as the model for the time until failure of a device. For example, the lifetime of a semiconductor chip might be modeled as an exponential random variable with a mean of 40,000 hours. The lack of memory property of the exponential distribution implies that the device does not wear out. That is, regardless of how long the device has been operating, the probability of a failure in the next 1000 hours is the same as the probability of a failure in the first 1000 hours of operation. The lifetime L of a device with failures caused by random shocks might be appropriately modeled as an exponential random variable.

However, the lifetime L of a device that suffers slow mechanical wear, such as bearing wear, is better modeled by a distribution such that $P(L < t + \Delta t \mid L > t)$ increases with t. Distributions such as the Weibull distribution are often used in practice to model the failure time of this type of device. The Weibull distribution is presented in a later section.

4.8 Erlang and Gamma Distributions

An exponential random variable describes the length until the first event is obtained in a Poisson process. A generalization of the exponential distribution is the length until r events occur in a Poisson process.

EXAMPLE 4.19 | Processor Failure

The failures of the central processor units of large computer systems are often modeled as a Poisson process. Typically, failures are not caused by components wearing out but by more random failures of the large number of semiconductor circuits in the units. Assume that the units that fail are immediately repaired and assume that the mean number of failures per hour is 0.0001. Let X denote the time until four failures occur in a system. Determine the probability that X exceeds 40,000 hours.

Let the random variable N denote the number of failures in 40,000 hours of operation. The time until four failures occur exceeds 40,000 hours if and only if the number of failures in 40,000 hours is three or less. Therefore,

$$P(X > 40{,}000) = P(N \leq 3)$$

The assumption that the failures follow a Poisson process implies that N has a Poisson distribution with

$$E(N) = \lambda T = 0.0001(40{,}000) = 4$$

failures per 40,000 hours. Therefore,

$$P(X > 40{,}000) = P(N \leq 3) = \sum_{k=0}^{3} \frac{e^{-4}4^k}{k!} = 0.433$$

The previous example can be generalized to show that if X is the time until the rth event in a Poisson process, then

$$P(X > x) = \sum_{k=0}^{r-1} \frac{e^{-\lambda x}(\lambda x)^k}{k!} \tag{4.17}$$

Because $P(X > x) = 1 - F(x)$, the probability density function of X equals the negative of the derivative of the right-hand side of the previous equation. After extensive algebraic simplification, the probability density function of X can be shown to equal

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!} \text{ for } x > 0 \text{ and } r = 1, 2, \ldots .$$

This probability density function defines an **Erlang random variable**. Clearly, an Erlang random variable with $r = 1$ is an exponential random variable.

It is convenient to generalize the Erlang distribution to allow r to assume any nonnegative value. Then the Erlang and some other common distributions become special cases of this generalized distribution. To accomplish this step, the factorial function $(r-1)!$ is generalized to apply to any nonnegative value of r, but the generalized function should still equal $(r-1)!$ when r is a positive integer.

Gamma Function

The **gamma function** is

$$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x}\, dx, \quad \text{for } r > 0 \tag{4.18}$$

It can be shown that the integral in the definition of $\Gamma(r)$ is finite. Furthermore, by using integration by parts, it can be shown that

$$\Gamma(r) = (r-1)\Gamma(r-1)$$

This result is left as an exercise. Therefore, if r is a positive integer (as in the Erlang distribution),

$$\Gamma(r) = (r-1)!$$

Also, $\Gamma(1) = 0! = 1$ and it can be shown that $\Gamma(1/2) = \pi^{1/2}$. Now the Erlang distribution can be generalized.

Gamma Distribution

The random variable X with probability density function

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}, \quad \text{for } x > 0 \tag{4.19}$$

is a **gamma random variable** with parameters $\lambda > 0$ and $r > 0$. If r is an integer, X has an Erlang distribution.

The parameters λ and r are often called the *scale* and *shape* parameters, respectively. However, one should check the definitions used in software packages. For example, some statistical software defines the scale parameter as $1/\lambda$. Sketches of the gamma distribution for several values of λ and r are shown in Figure 4.21. Many different shapes can be generated from changes to the parameters.

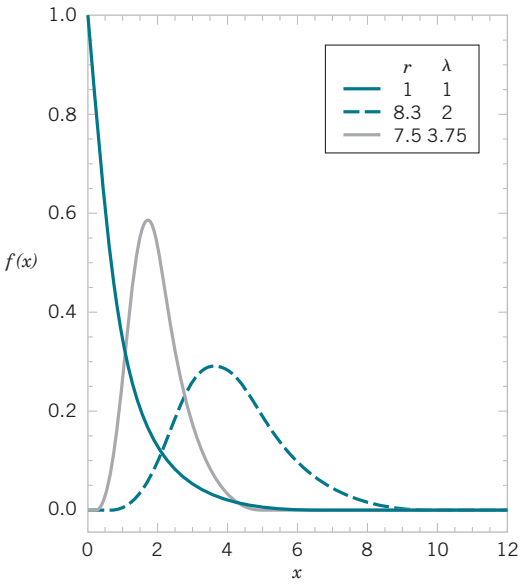


FIGURE 4.21
Gamma probability density functions for selected values of λ and *r*.

Also, the change of variable $u = \lambda x$ and the definition of the gamma function can be used to show that the probability density function integrates to 1.

For the special case when r is an integer and the value of r is not large, Equation 4.17 can be applied to calculate probabilities for a gamma random variable. However, in general, the integral of the gamma probability density function is difficult to evaluate so computer software is used to determine probabilities.

Recall that for an exponential distribution with parameter λ, the mean and variance are $1/\lambda$ and $1/\lambda^2$, respectively. An Erlang random variable is the time until the rth event in a Poisson process and the time between events are independent. Therefore, it is plausible that the mean and variance of a gamma random variable multiply the exponential results by r. This motivates the following conclusions. Repeated integration by parts can be used to derive these, but the details are lengthy and omitted.

Mean and Variance

If X is a **gamma random variable** with parameters λ and r,

$$\mu = E(X) = r/\lambda \quad \text{and} \quad \sigma^2 = V(X) = r/\lambda^2$$

EXAMPLE 4.20

The time to prepare a slide for high-throughput genomics is a Poisson process with a mean of two hours per slide. What is the probability that 10 slides require more than 25 hours to prepare?

Let X denote the time to prepare 10 slides. Because of the assumption of a Poisson process, X has a gamma distribution with $\lambda = 1/2$, $r = 10$, and the requested probability is $P(X > 25)$. The probability can be obtained from software that provides cumulative Poisson probabilities or gamma probabilities. For the cumulative Poisson probabilities, we use the method in Example 4.19 to obtain

$$P(X > 25) = \sum_{k=0}^{9} \frac{e^{-12.5}(12.5)^k}{k!}$$

In software we set the mean = 12.5 and the input = 9 to obtain $P(X > 25) = 0.2014$.

What are the mean and standard deviation of the time to prepare 10 slides? The mean time is

$$E(X) = r/\lambda = 10/0.5 = 20$$

The variance of time is

$$V(X) = r/\lambda^2 = 10/0.5^2 = 40$$

so that the standard deviation is $40^{1/2} = 6.32$ hours.

The slides will be completed by what length of time with probability equal to 0.95? The question asks for x such that

$$P(X \leq x) = 0.95$$

where X is gamma with $\lambda = 0.5$ and $r = 10$. In software, we use the gamma inverse cumulative probability function and set the shape parameter to 10, the scale parameter to 0.5, and the probability to 0.95. The solution is

$$P(X \leq 31.41) = 0.95$$

Practical Interpretation: Based on this result, a schedule that allows 31.41 hours to prepare 10 slides should be met 95% of the time.

Furthermore, the **chi-square distribution** is a special case of the gamma distribution in which $\lambda = 1/2$ and r equals one of the values 1/2, 1, 3/2, 2, ... This distribution is discussed and used extensively in interval estimation and tests of hypotheses in subsequent chapters.

4.9 Weibull Distribution

As mentioned previously, the Weibull distribution is often used to model the time until failure of many different physical systems. The parameters in the distribution provide a great deal of flexibility to model systems in which the number of failures increases with time (bearing wear), decreases with time (some semiconductors), or remains constant with time (failures caused by external shocks to the system).

Weibull Distribution

The random variable X with probability density function

$$f(x) = \frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\delta}\right)^{\beta}\right], \quad \text{for } x > 0 \tag{4.20}$$

is a **Weibull random variable** with scale parameter $\delta > 0$ and shape parameter $\beta > 0$.

The graphs of selected probability density functions in Figure 4.22 illustrate the flexibility of the Weibull distribution. By inspecting the probability density function, we can see that when $\beta = 1$, the Weibull distribution is identical to the exponential distribution. Also, the **Raleigh distribution** is a special case when the shape parameter is 2.

The cumulative distribution function is often used to compute probabilities. The following result can be obtained.

Cumulative Distribution Function

If X has a Weibull distribution with parameters δ and β, then the cumulative distribution function of X is

$$F(x) = 1 - e^{-\left(\frac{x}{\delta}\right)^{\beta}}$$

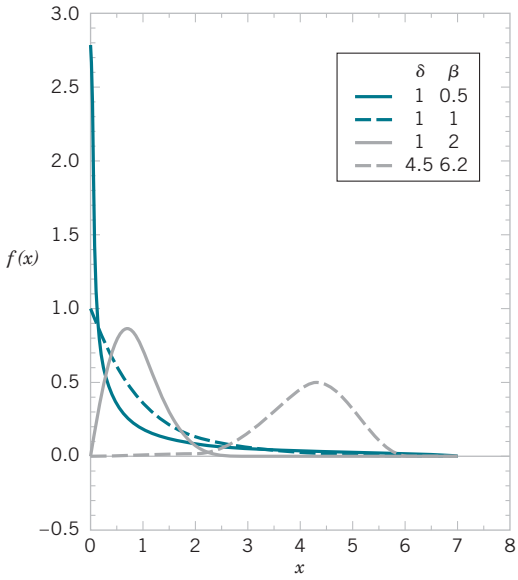


FIGURE 4.22
Weibull probability density functions for selected values of δ and β.

Also, the following results can be obtained.

Mean and Variance

If X has a Weibull distribution with parameters δ and β,

$$\mu = E(X) = \delta\Gamma\left(1+\frac{1}{\beta}\right) \quad \text{and} \quad \sigma^2 = V(X) = \delta^2\Gamma\left(1+\frac{2}{\beta}\right) - \delta^2\left[\Gamma\left(1+\frac{1}{\beta}\right)\right]^2 \tag{4.21}$$

EXAMPLE 4.21 | Bearing Wear

The time to failure (in hours) of a bearing in a mechanical shaft is satisfactorily modeled as a Weibull random variable with $\beta = 1/2$ and $\delta = 5000$ hours. Determine the mean time until failure.

From the expression for the mean,

$$E(X) = 5000\Gamma[1 + (1/2)] = 5000\Gamma[1.5] = 5000 \times 0.5\sqrt{\pi}$$
$$= 4431.1 \text{ hours}$$

Determine the probability that a bearing lasts at least 6000 hours. Now,

$$P(X > 6000) = 1 - F(6000) = \exp\left[-\left(\frac{6000}{5000}\right)^2\right]$$
$$= e^{-1.44} = 0.237$$

Practical Interpretation: Consequently, only 23.7% of all bearings last at least 6000 hours.

4.10 Lognormal Distribution

Variables in a system sometimes follow an exponential relationship as $x = \exp(w)$. If the exponent is a random variable W, then $X = \exp(W)$ is a random variable with a distribution of interest. An important special case occurs when W has a normal distribution. In that case, the distribution of X is called a **lognormal distribution**. The name follows from the transformation $\ln(X) = W$. That is, the natural logarithm of X is normally distributed.

Probabilities for X are obtained from the transform of the normal distribution. Suppose that W is normally distributed with mean θ and variance ω^2; then the cumulative distribution function for X is

$$F(x) = P[X \le x] = P[\exp(W) \le x] = P[W \le \ln(x)]$$
$$= P\left[Z \le \frac{\ln(x) - \theta}{\omega}\right] = \Phi\left[\frac{\ln(x) - \theta}{\omega}\right]$$

for $x > 0$, where Z is a standard normal random variable and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Therefore, Appendix Table III can be used to determine the probability. Also, $F(x) = 0$ for $x \le 0$.

The probability density function of X can be obtained from the derivative of $F(x)$. Because $\Phi(\cdot)$ is the integral of the standard normal density function, the fundamental theorem of calculus is used to calculate the derivative. Furthermore, from the probability density function, the mean and variance of X can be derived. The details are omitted, but a summary of results follows.

Lognormal Distribution

Let W have a normal distribution with mean θ and variance ω^2; then $X = \exp(W)$ is a **lognormal random variable** with probability density function

$$f(x) = \frac{1}{x\omega\sqrt{2\pi}} \exp\left[-\frac{(\ln(x) - \theta)^2}{2\omega^2}\right] \qquad 0 < x < \infty$$

The mean and variance of X are

$$E(X) = e^{\theta+\omega^2/2} \quad \text{and} \quad V(X) = e^{2\theta+\omega^2}(e^{\omega^2} - 1) \tag{4.22}$$

The parameters of a lognormal distribution are θ and ω^2, but these are the mean and variance of the normal random variable W. The mean and variance of X are the functions of these parameters shown in Equation 4.22. Figure 4.23 illustrates lognormal distributions for selected values of the parameters.

The lifetime of a product that degrades over time is often modeled by a lognormal random variable. For example, this is a common distribution for the lifetime of a semiconductor laser. A Weibull distribution can also be used in this type of application, and with an appropriate choice

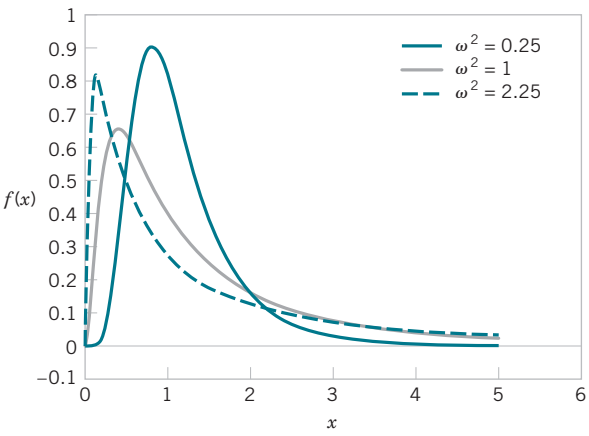


FIGURE 4.23
Lognormal probability density functions with $\theta = 0$ for selected values of ω^2.

for parameters, it can approximate a selected lognormal distribution. However, a lognormal distribution is derived from a simple exponential function of a normal random variable, so it is easy to understand and easy to evaluate probabilities.

EXAMPLE 4.22 | Semiconductor Laser

The lifetime (in hours) of a semiconductor laser has a lognormal distribution with $\theta = 10$ and $\omega = 1.5$. What is the probability that the lifetime exceeds 10,000 hours?

From the cumulative distribution function for X,

$$\begin{aligned} P(X > 10{,}000) &= 1 - P[\exp(W) \le 10{,}000] \\ &= 1 - P[W \le \ln(10{,}000)] \\ &= 1 - \Phi\left(\frac{\ln(10{,}000) - 10}{1.5}\right) = 1 - \Phi(-0.52) \\ &= 1 - 0.30 = 0.70 \end{aligned}$$

What lifetime is exceeded by 99% of lasers? The question is to determine x such that $P(X > x) = 0.99$. Therefore,

$$\begin{aligned} P(X > x) &= P[\exp(W) > x] = P[W > \ln(x)] \\ &= 1 - \Phi\left(\frac{\ln(x) - 10}{1.5}\right) = 0.99 \end{aligned}$$

From Appendix Table III, $1 - \Phi(z) = 0.99$ when $z = -2.33$. Therefore,

$$\frac{\ln(x) - 10}{1.5} = -2.33 \quad \text{and} \quad x = \exp(6.505) = 668.48 \text{ hours}$$

Determine the mean and standard deviation of lifetime. Now,

$$\begin{aligned} E(X) &= e^{\theta + \omega^2/2} = e^{(10 + 1.125)} = 67{,}846.3 \\ V(X) &= e^{2\theta + \omega^2}(e^{\omega^2} - 1) = e^{(20 + 2.25)}(e^{2.25} - 1) \\ &= 39{,}070{,}059{,}886.6 \end{aligned}$$

hours squared, so the standard deviation of X is 197,661.5 hours.

Practical Interpretation: The standard deviation of a lognormal random variable can be large relative to the mean.

4.11 Beta Distribution

A continuous distribution that is flexble but bounded over a finite range is useful for probability models. The proportion of solar radiation absorbed by a material or the proportion (of the maximum time) required to complete a task in a project are examples of continuous random variables over the interval [0, 1].

The random variable X with probability density function

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, x^{\alpha - 1}(1 - x)^{\beta - 1}, \quad \text{for } x \text{ in } [0, 1]$$

is a **beta random variable** with parameters $\alpha > 0$ and $\beta > 0$.

The shape parameters α and β allow the probability density function to assume many different shapes. Figure 4.24 provides some examples. If $\alpha = \beta$, the distribution is symmetric about $x = 0.5$, and if $\alpha = \beta = 1$, the beta distribution equals a continuous uniform distribution. Figure 4.24 illustrates that other parameter choices generate nonsymmetric distributions.

In general, there is not a closed-form expression for the cumulative distribution function, and probabilities for beta random variables need to be computed numerically. The exercises provide some special cases in which the probability density function is more easily handled.

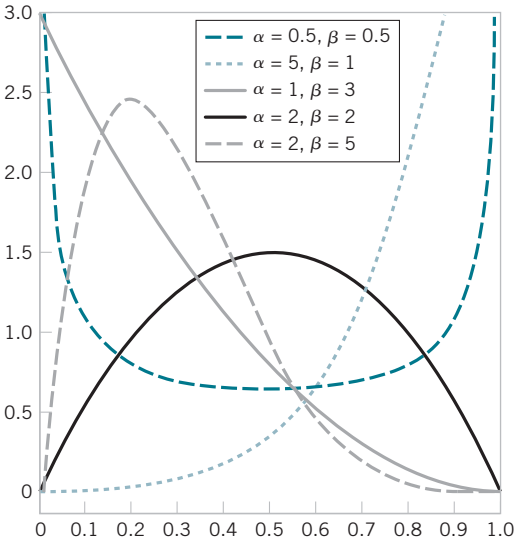


FIGURE 4.24
Beta probability density functions for selected values of the parameters α and β.

EXAMPLE 4.23

The service of a constant-velocity joint in an automobile requires disassembly, boot replacement, and assembly. Suppose that the proportion of the total service time for disassembly follows a beta distribution with $\alpha = 2.5$ and $\beta = 1$. What is the probability that a disassembly proportion exceeds 0.7?

Let X denote the proportion of service time for disassembly. The probability is

$$P(X > 0.7) = \int_{0.7}^{1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

$$= \int_{0.7}^{1} \frac{\Gamma(3.5)}{\Gamma(2.5)\,\Gamma(1)} x^{1.5} = \frac{2.5(1.5)(0.5)\sqrt{\pi}}{(1.5)(0.5)\sqrt{\pi}} \frac{x^{2.5}}{2.5}\Bigg|_{0.7}^{1}$$

$$= 1 - 0.7^{2.5} = 0.59$$

If $\alpha > 1$ and $\beta > 1$, the mode (peak of the density) is in the interior of [0, 1] and equals

$$\text{mode} = \frac{\alpha - 1}{\alpha + \beta - 2}$$

This expression is useful to relate the peak of the probability density function to the parameters. Suppose that the proportion of time to complete one task among several follows a beta distribution with $\alpha = 2.5$ and $\beta = 1$. The mode of this distribution is $(2.5 - 1)/(3.5 - 2) = 1$. The mean and variance of a beta distribution can be obtained from the integrals, but the details are left as an exercise.

Also, although a beta random variable X is defined over the interval [0, 1], a generalized beta random variable W, defined over the finite interval $[a, b]$, can be constructed from $W = a + (b - a)X$.

Mean and Variance

If X has a beta distribution with parameters α and β,

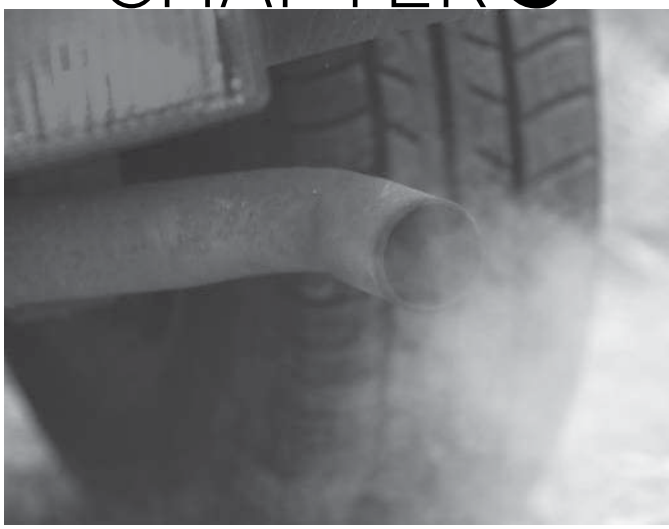
$$\mu = E(X) = \frac{\alpha}{\alpha + \beta} \quad \text{and} \quad \sigma^2 = V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Important Terms and Concepts

Beta random variable
Chi-squared distribution
Continuity correction
Continuous random variable
Continuous uniform random variable
Cumulative distribution function
Erlang random variable
Expected value of a function of a random variable
Exponential random variable
Gamma function
Gamma random variable
Gaussian distribution
Lack of memory property—continuous random variable
Lognormal random variable
Mean—continuous random variable
Mean—function of a continuous random variable
Normal approximation to binomial and Poisson probabilities
Normal random variable
Poisson process
Probability density function
Probability distribution—continuous random variable
Raleigh distribution
Standard deviation—continuous random variable
Standard normal random variable
Standardizing
Variance—continuous random variable
Weibull random variable

CHAPTER 5

Joint Probability Distributions



© Ralph125 / iStockphoto

LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Use joint probability mass functions and joint probability density functions to calculate probabilities and calculate marginal probability distributions from joint probability distributions
2. Calculate conditional probability distributions from joint probability distributions and assess independence of random variables
3. Interpret and calculate covariances and correlations between random variables
4. Use the multinomial distribution to determine probabilities and understand properties of a bivariate normal distribution and be able to draw contour plots for the probability density function
5. Calculate means and variances for linear functions of random variables and calculate probabilities for linear functions of normally distributed random variables
6. Determine the distribution of a general function of a random variable
7. Calculate the moment-generating functions and use the functions to determine moments and distributions.

CHAPTER OUTLINE

5.1 Joint Probability Distributions for Two Random Variables

5.2 Conditional Probability Distributions and Independence

5.3 Joint Probability Distributions for More Than Two Random Variables

5.4 Covariance and Correlation

5.5 Common Joint Distributions

5.5.1 Multinomial Probability Distribution

5.5.2 Bivariate Normal Distribution

5.6 Linear Functions of Random Variables

5.7 General Functions of Random Variables

5.8 Moment-Generating Functions

Air-quality monitoring stations are maintained throughout Maricopa County, Arizona, and the Phoenix metropolitan area. Measurements for particulate matter and ozone are measured hourly. Particulate matter (known as PM10) is a measure (in $\mu g/m^3$) of solid and liquid particles in the air with diameters less than 10 micrometers. Ozone is a colorless gas with molecules comprised of three oxygen atoms that make it very reactive. Ozone is formed in a complex reaction from heat, sunlight, and other pollutants, especially volatile organic compounds. The U.S. Environmental Protection Agency sets limits for both PM10 and ozone. For example, the limit for ozone is 0.075 ppm. The probability that a day in Phoenix exceeds the limits for PM10 and ozone is important for compliance and remedial actions with the county and city. But this might be more involved than the product of the probabilities for each pollutant separately. It might be that days with high PM10 measurements also tend to have ozone values. That is, the measurements might not be independent, so the joint relationship between these measurements becomes important. The study of probability distributions for more than one random variable is the focus of this chapter and the air-quality data is just one illustration of the ubiquitous need to study variables jointly.

5.1 Joint Probability Distributions for Two Random Variables

In Chapters 3 and 4, you studied probability distributions for a single random variable. However, it is often useful to have more than one random variable defined in a random experiment. For example, the continuous random variable X can denote the length of one dimension of an injection-molded part, and the continuous random variable Y might denote the length of another dimension. We might be interested in probabilities that can be expressed in terms of both X and Y. If the specifications for X and Y are (2.95 to 3.05) and (7.60 to 7.80) millimeters, respectively, we might be interested in the probability that a part satisfies both specifications; that is, $P(2.95 < X < 3.05 \text{ and } 7.60 < Y < 7.80)$.

Because the two random variables are measurements from the same part, small disturbances in the injection-molding process, such as pressure and temperature variations, might be more likely to generate values for X and Y in specific regions of two-dimensional space. For example, a small pressure increase might generate parts such that both X and Y are greater than their respective targets, and a small pressure decrease might generate parts such that X and Y are both less than their respective targets. Therefore, based on pressure variations, we expect that the probability of a part with X much greater than its target and Y much less than its target is small.

In general, if X and Y are two random variables, the probability distribution that defines their simultaneous behavior is called a **joint probability distribution**. In this chapter, we investigate some important properties of these joint distributions. For simplicity, we begin by considering random experiments in which only two random variables are studied. In later sections, we generalize the presentation to the joint probability distribution of more than two random variables.

EXAMPLE 5.1 | Mobile Response Time

The response time is the speed of page downloads and it is critical for a mobile Web site. As the response time increases, customers become more frustrated and potentially abandon the site for a competitive one. Let X denote the number of bars of service, and let Y denote the response time (to the nearest second) for a particular user and site.

By specifying the probability of each of the points in Figure 5.1, we specify the joint probability distribution of X and Y. Similarly to an individual random variable, we define the range of the random variables (X, Y) to be the set of points (x, y) in two-dimensional space for which the probability that $X = x$ and $Y = y$ is positive.

y = Response Time (nearest second)	x = Number of Bars of Signal Strength		
	1	2	3
4	0.15	0.1	0.05
3	0.02	0.1	0.05
2	0.02	0.03	0.2
1	0.01	0.02	0.25

FIGURE 5.1

Joint probability distribution of X and Y in Example 5.1.

Joint Probability Mass Function If X and Y are discrete random variables, the joint probability distribution of X and Y is a description of the set of points (x, y) in the range of (X, Y) along with the probability of each point. Also, $P(X = x \text{ and } Y = y)$ is usually written as $P(X = x, Y = y)$. The joint probability distribution of two random variables is sometimes referred to as the **bivariate probability distribution** or **bivariate distribution** of the random variables. One way to describe the joint probability distribution of two discrete random variables is through a joint probability mass function $f(x, y) = P(X = x, Y = y)$.

Joint Probability Mass Function

The **joint probability mass function** of the discrete random variables X and Y, denoted as $f_{XY}(x, y)$, satisfies

$$(1)\ f_{XY}(x, y) \geq 0$$
$$(2)\ \sum_X \sum_Y f_{XY}(x, y) = 1$$
$$(3)\ f_{XY}(x, y) = P(X = x, Y = y) \qquad (5.1)$$

Just as the probability mass function of a single random variable X is assumed to be zero at all values outside the range of X, so the joint probability mass function of X and Y is assumed to be zero at values for which a probability is not specified.

Joint Probability Density Function The joint probability distribution of two continuous random variables X and Y can be specified by providing a method for calculating the probability that X and Y assume a value in any region R of two-dimensional space. Analogous to the probability density function of a single continuous random variable, a **joint probability density function** can be defined over two-dimensional space. The double integral of $f_{XY}(x, y)$ over a region R provides the probability that (X, Y) assumes a value in R. This integral can be interpreted as the volume under the surface $f_{XY}(x, y)$ over the region R. Typically, $f_{XY}(x, y)$ is defined over all of two-dimensional space by assuming that $f_{XY}(x, y) = 0$ for all points for which $f_{XY}(x, y)$ is not specified.

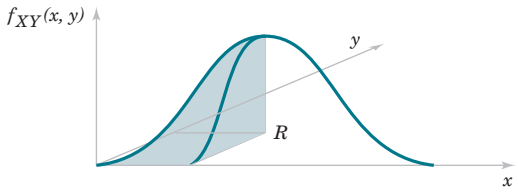


FIGURE 5.2

Joint probability density function for random variables X and Y. Probability that (X, Y) is in the region R is determined by the volume of $f_{XY}(x, y)$ over the region R.

A joint probability density function for X and Y is shown in Figure 5.2. The probability that (X, Y) assumes a value in the region R equals the volume of the shaded region in Figure 5.2. In this manner, a joint probability density function is used to determine probabilities for X and Y.

Joint Probability Density Function

A **joint probability density function** for the continuous random variables X and Y, denoted as $f_{XY}(x, y)$, satisfies the following properties:

(1) $f_{XY}(x, y) \geq 0$ for all x, y

(2) $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)\, dx\, dy = 1$$

(3) For any region R of two-dimensional space,

$$P((X, Y) \in R) = \iint_R f_{XY}(x, y)\, dx\, dy \tag{5.2}$$

At the start of this chapter, the lengths of different dimensions of an injection-molded part were presented as an example of two random variables. However, because the measurements are from the same part, the random variables are typically not independent. If the specifications for X and Y are [2.95, 3.05] and [7.60, 7.80] millimeters, respectively, we might be interested in the probability that a part satisfies both specifications; that is, $P(2.95 < X < 3.05, 7.60 < Y < 7.80)$. Suppose that $f_{XY}(x, y)$ is shown in Figure 5.3. The required probability is the volume of $f_{XY}(x, y)$ within the specifications. Often a probability such as this must be determined from a numerical integration.

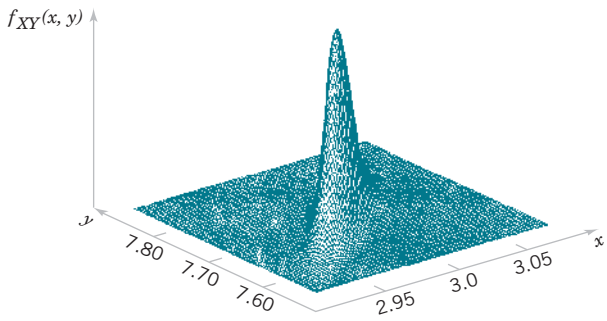


FIGURE 5.3

Joint probability density function for the lengths of different dimensions of an injection-molded part.

EXAMPLE 5.2 | Server Access Time

Let the random variable X denote the time until a computer server connects to your machine (in milliseconds), and let Y denote the time until the server authorizes you as a valid user (in milliseconds). Each of these random variables measures the wait from a common starting time and $X < Y$. Assume that the joint probability density function for X and Y is

$$f_{XY}(x, y) = 6 \times 10^{-6} \exp(-0.001x - 0.002y) \quad \text{for} \quad x < y$$

Reasonable assumptions can be used to develop such a distribution, but for now, our focus is on only the joint probability density function.

The region with nonzero probability is shaded in Figure 5.4. The property that this joint probability density function integrates to 1 can be verified by the integral of $f_{XY}(x, y)$ over this region as follows:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dy\,dx = \int_{0}^{\infty}\left(\int_{x}^{\infty} 6\times 10^{-6} e^{-0.001x-0.002y}\,dy\right)dx$$

$$= 6\times 10^{-6}\int_{0}^{\infty}\left(\int_{x}^{\infty} e^{-0.002y}\,dy\right)e^{-0.001x}\,dx$$

$$= 6\times 10^{-6}\int_{0}^{\infty}\left(\frac{e^{-0.002x}}{0.002}\right)e^{-0.001x}\,dx$$

$$= 0.003\left(\int_{0}^{\infty} e^{-0.003x}\,dx\right) = 0.003\left(\frac{1}{0.003}\right) = 1$$

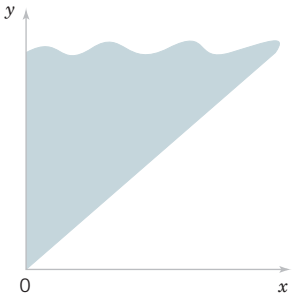


FIGURE 5.4

The joint probability density function of X and Y is nonzero over the shaded region.

The probability that $X < 1000$ and $Y < 2000$ is determined as the integral over the darkly shaded region in Figure 5.5.

$$P(X \le 1000, Y \le 2000) = \int_{0}^{1000}\int_{x}^{2000} f_{XY}(x,y)\,dy\,dx$$

$$= 6\times 10^{-6}\int_{0}^{1000}\left(\int_{x}^{2000} e^{-0.002y}\,dy\right)e^{-0.001x}\,dx$$

$$= 6\times 10^{-6}\int_{0}^{1000}\left(\frac{e^{-0.002x}-e^{-4}}{0.002}\right)e^{-0.001x}\,dx$$

$$= 0.003\int_{0}^{1000} e^{-0.003x} - e^{-4}e^{-0.001x}\,dx$$

$$= 0.003\left[\left(\frac{1-e^{-3}}{0.003}\right) - e^{-4}\left(\frac{1-e^{-1}}{0.001}\right)\right]$$

$$= 0.003\,(316.738 - 11.578) = 0.915$$

Practical Interpretation: A joint probability density function enables probabilities for two (or more) random variables to be calculated as in these examples.

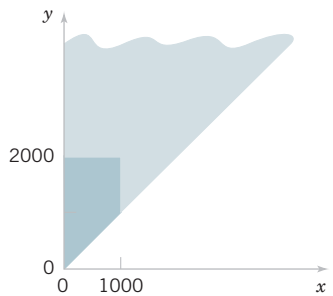


FIGURE 5.5

Region of integration for the probability that $X < 1000$ and $Y < 2000$ is darkly shaded.

Marginal Probability Distributions If more than one random variable is defined in a random experiment, it is important to distinguish between the joint probability distribution of X and Y and the probability distribution of each variable individually. The individual probability distribution of a random variable is referred to as its **marginal probability distribution**.

The marginal probability distribution of X can be determined from the joint probability distribution of X and other random variables. For example, consider discrete random variables

X and Y. To determine $P(X = x)$, we sum $P(X = x, Y = y)$ over all points in the range of (X, Y) for which $X = x$. Subscripts on the probability mass functions distinguish between the random variables.

EXAMPLE 5.3 | Marginal Distribution

The joint probability distribution of X and Y in Figure 5.1 can be used to find the marginal probability distribution of X. For example,

$$f_X(3) = P(X = 3) = P(X = 3, Y = 1) + P(X = 3, Y = 2) + P(X = 3, Y = 3) + P(X = 3, Y = 4)$$
$$= 0.25 + 0.2 + 0.05 + 0.05 = 0.55$$

The marginal probability distribution for X is found by summing the probabilities in each column whereas the marginal probability distribution for Y is found by summing the probabilities in each row. The results are shown in Figure 5.6.

	x = Number of Bars of Signal Strength			
y = Response Time (nearest second)	1	2	3	Marginal Probability Distribution of Y
4	0.15	0.1	0.05	**0.3**
3	0.02	0.1	0.05	**0.17**
2	0.02	0.03	0.2	**0.25**
1	0.01	0.02	0.25	**0.28**
	0.2	**0.25**	**0.55**	
	Marginal Probability Distribution of X			

FIGURE 5.6

Marginal probability distributions of X and Y from Figure 5.1.

For continuous random variables, an analogous approach is used to determine marginal probability distributions. In the continuous case, an integral replaces the sum.

Marginal Probability Density Function

If the joint probability density function of random variables X and Y is $f_{XY}(x, y)$, the **marginal probability density functions** of X and Y are

$$f_X(x) = \int f_{XY}(x, y)\, dy \quad \text{and} \quad f_Y(y) = \int f_{XY}(x, y)\, dx \tag{5.3}$$

where the first integral is over all points in the range of (X, Y) for which $X = x$ and the second integral is over all points in the range of (X, Y) for which $Y = y$.

A probability for only one random variable, say, for example, $P(a < X < b)$, can be found from the marginal probability distribution of X or from the integral of the joint probability distribution of X and Y as

$$P(a < X < b) = \int_a^b f_X(x)\, dx = \int_a^b \left[\int_{-\infty}^{\infty} f_{XY}(x, y)\, dy \right] dx = \int_a^b \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy dx$$

EXAMPLE 5.4 | Server Access Time

For the random variables that denote times in Example 5.2, calculate the probability that Y exceeds 2000 milliseconds.

This probability is determined as the integral of $f_{XY}(x, y)$ over the darkly shaded region in Figure 5.7. The region is partitioned into two parts and different limits of integration are determined for each part.

$$P(Y > 2000) = \int_0^{2000}\left(\int_{2000}^{\infty} 6\times 10^{-6} e^{-0.001x-0.002y}\, dy\right) dx + \int_{2000}^{\infty}\left(\int_{x}^{\infty} 6\times 10^{-6} e^{-0.001x-0.002y}\, dy\right) dx$$

The first integral is

$$6\times 10^{-6}\int_0^{2000}\left(\frac{e^{-0.002y}}{-0.002}\Big|_{2000}^{\infty}\right)e^{-0.001x}\, dx = \frac{6\times 10^{-6}}{0.002}e^{-4}$$

$$\times \int_0^{2000} e^{-0.001x}\, dx = \frac{6\times 10^{-6}}{0.002}e^{-4}\left(\frac{1-e^{-2}}{0.001}\right) = 0.0475$$

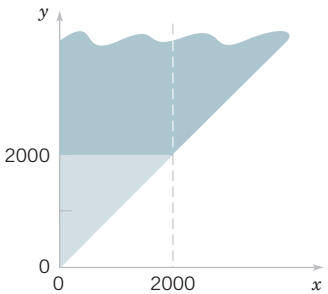


FIGURE 5.7
Region of integration for the probability that $Y > 2000$ is darkly shaded, and it is partitioned into two regions with $x < 2000$ and $x > 2000$.

The second integral is

$$6\times 10^{-6}\int_{2000}^{\infty}\left(\frac{e^{-0.002y}}{-0.002}\Big|_{x}^{\infty}\right)e^{-0.001x}\, dx = \frac{6\times 10^{-6}}{0.002}$$

$$\times \int_{2000}^{\infty} e^{-0.003x}\, dx = \frac{6\times 10^{-6}}{0.002}\left(\frac{e^{-6}}{0.003}\right) = 0.0025$$

Therefore,

$$P(Y > 2000) = 0.0475 + 0.0025 = 0.05$$

Alternatively, the probability can be calculated from the marginal probability distribution of Y as follows. For $y > 0$,

$$f_Y(y) = \int_0^{y} 6\times 10^{-6} e^{-0.001x-0.002y}\, dx = 6\times 10^{-6} e^{-0.002y}$$

$$\times \int_0^{y} e^{-0.001x}\, dx = 6\times 10^{-6} e^{-0.002y}\left(\frac{e^{-0.001x}}{-0.001}\Big|_0^{y}\right)$$

$$= 6\times 10^{-6} e^{-0.002y}\left(\frac{1-e^{-0.001y}}{0.001}\right)$$

$$= 6\times 10^{-3} e^{-0.002y}(1-e^{-0.001y}) \quad \text{for } y > 0$$

We have obtained the marginal probability density function of Y. Now,

$$P(Y > 2000) = 6\times 10^{-3}\int_{2000}^{\infty} e^{-0.002y}(1-e^{-0.001y})\, dy$$

$$= 6\times 10^{-3}\left[\left(\frac{e^{-0.002y}}{-0.002}\Big|_{2000}^{\infty}\right) - \left(\frac{e^{-0.003y}}{-0.003}\Big|_{2000}^{\infty}\right)\right]$$

$$= 6\times 10^{-3}\left[\frac{e^{-4}}{0.002} - \frac{e^{-6}}{0.003}\right] = 0.05$$

Also, $E(X)$ and $V(X)$ can be determined from the marginal probability distribution of X in the usual manner or from the joint probability distribution of X and Y as follows.

Mean and Variance from a Joint Distribution

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x, y)dydx$$

and

$$V(X) = \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)^2 f_{X,Y}(x, y)dydx \qquad (5.4)$$

In Figure 5.6, the marginal probability distributions of X and Y are used to obtain the means as

$$E(X) = 1(0.2) + 2(0.25) + 3(0.55) = 2.35$$

$$E(Y) = 1(0.28) + 2(0.25) + 3(0.17) + 4(0.3) = 2.49$$

5.2 Conditional Probability Distributions and Independence

When two random variables are defined in a random experiment, knowledge of one can change the probabilities that we associate with the values of the other. Recall that in Example 5.1, X denotes the number of bars of service and Y denotes the response time. One expects the probability $Y = 1$ to be greater at $X = 3$ bars than at $X = 1$ bar. From the notation for conditional probability in Chapter 2, we can write such conditional probabilities as $P(Y = 1 \mid X = 3)$ and $P(Y = 1 \mid X = 1)$. Consequently, the random variables X and Y are expected to be dependent. Knowledge of the value obtained for X changes the probabilities associated with the values of Y.

Recall that the definition of conditional probability for events A and B is $P(B \mid A) = P(A \cap B)/P(A)$. This definition can be applied with the event A defined to be $X = x$ and event B defined to be $Y = y$.

EXAMPLE 5.5 | Conditional Probabilities for Mobile Response Time

For Example 5.1, X and Y denote the number of bars of signal strength and response time, respectively. Then,

$$P(Y = 1 \mid X = 3) = \frac{P(X = 3, Y = 1)}{P(X = 3)} = \frac{f_{XY}(3, 1)}{f_X(3)}$$
$$= \frac{0.25}{0.55} = 0.454$$

The probability that $Y = 2$ given that $X = 3$ is

$$P(Y = 2 \mid X = 3) = \frac{P(X = 3, Y = 2)}{P(X = 3)} = \frac{f_{XY}(3, 2)}{f_X(3)}$$
$$= \frac{0.2}{0.55} = 0.364$$

Further work shows that

$$P(Y = 3 \mid X = 3) = 0.091$$

and

$$P(Y = 4 \mid X = 3) = 0.091$$

Note that $P(Y = 1 \mid X = 3) + P(Y = 2 \mid X = 3) + P(Y = 3 \mid X = 3) + P(Y = 4 \mid X = 3) = 1$. This set of probabilities defines the conditional probability distribution of Y given that $X = 3$.

Example 5.5 illustrates that the conditional probabilities for Y given that $X = x$ can be thought of as a new probability distribution called the **conditional probability mass function** for Y given $X = x$. For Example 5.5, the conditional probability mass function for Y given that $X = 3$ consists of the four probabilities $f_{Y|3}(1) = 0.454, f_{Y|3}(2) = 0.364, f_{Y|3}(3) = 0.091, f_{Y|3}(4) = 0.091$.

The following definition applies these concepts to continuous random variables.

Conditional Probability Density Function

Given continuous random variables X and Y with joint probability density function $f_{XY}(x, y)$, the **conditional probability density function** of Y given $X = x$ is

$$f_{Y|x}(y) = \frac{f_{XY}(x, y)}{f_X(x)} \quad \text{for} \quad f_X(x) > 0 \tag{5.5}$$

The conditional probability density function provides the conditional probabilities for the values of Y given that $X = x$.

Because the conditional probability density function $f_{Y|x}(y)$ is a probability density function for all y in R_x, the following properties are satisfied:

(1) $f_{Y|x}(y) \geq 0$

(2) $\int f_{Y|x}(y)\, dy = 1$

(3) $P(Y \in B \mid X = x) = \int_B f_{Y|x}(y)\, dy$ for any set B in the range of Y (5.6)

It is important to state the region in which a joint, marginal, or conditional probability density function is not zero. The following example illustrates this.

EXAMPLE 5.6 | Conditional Probability

For the random variables that denote times in Example 5.2, determine the conditional probability density function for Y given that $X = x$.

First, the marginal density function of x is determined. For $x > 0$,

$$f_X(x) = \int_x 6 \times 10^{-6} e^{-0.001x - 0.002y}\, dy$$

$$= 6 \times 10^{-6} e^{-0.001x} \left(\frac{e^{-0.002y}}{-0.002} \Big|_x^\infty \right)$$

$$= 6 \times 10^{-6} e^{-0.001x} \left(\frac{e^{-0.002x}}{0.002} \right)$$

$$= 0.003 e^{-0.003x} \quad \text{for} \quad x > 0$$

This is an exponential distribution with $\lambda = 0.003$. Now for $0 < x$ and $x < y$, the conditional probability density function is

$$f_{Y|x}(y) = \frac{f_{XY}(x, y)}{f_x(x)} = \frac{6 \times 10^{-6} e^{-0.001x - 0.002y}}{0.003 e^{-0.003x}}$$

$$= 0.002 e^{0.002x - 0.002y} \quad \text{for} \quad 0 < x \text{ and } x < y$$

The conditional probability density function of Y, given that $X = 1500$, is nonzero on the solid line in Figure 5.8.

Determine the probability that Y exceeds 2000, given that $x = 1500$. That is, determine $P(Y > 2000 \mid X = 1500)$.

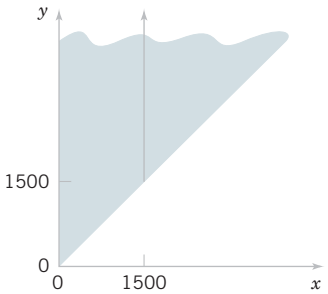


FIGURE 5.8

The conditional probability density function for Y, given that $x = 1500$, is nonzero over the solid line.

The conditional probability density function is integrated as follows:

$$P(Y > 2000 \mid X = 1500) = \int_{2000}^{\infty} f_{Y|1500}(y)\, dy$$

$$= \int_{2000}^{\infty} 0.002 e^{0.002(1500) - 0.002y}\, dy$$

$$= 0.002 e^3 \left(\frac{e^{-0.002y}}{-0.002} \Big|_{2000}^\infty \right)$$

$$= 0.002 e^3 \left(\frac{e^{-4}}{0.002} \right) = 0.368$$

Conditional Mean and Variance

The **conditional mean** of Y given $X = x$, denoted as $E(Y \mid x)$ or $\mu_{Y|x}$, is

$$E(Y \mid x) = \int_y y f_{Y|x}(y) \tag{5.7}$$

and the **conditional variance** of Y given $X = x$, denoted as $V(Y \mid x)$ or $\sigma^2_{Y|x}$, is

$$V(Y \mid x) = \int_y (y - \mu_{Y|x})^2 f_{Y|x}(y) = \int_y y^2 f_{Y|x}(y) - \mu^2_{Y|x}$$

EXAMPLE 5.7 | Conditional Mean and Variance

For the random variables that denote times in Example 5.2, determine the conditional mean for Y given that $x = 1500$.

The conditional probability density function for Y was determined in Example 5.6. Because $f_{Y|1500}(y)$ is nonzero for $y > 1500$,

$$E(Y \mid X = 1500) = \int_{1500}^{\infty} y(0.002e^{0.002(1500)-0.002y})\, dy$$

$$= 0.002e^3 \int_{1500}^{\infty} ye^{-0.002y}\, dy$$

Integrate by parts as follows:

$$\int_{1500}^{\infty} ye^{-0.002y}\, dy = y\frac{e^{-0.002y}}{-0.002}\Big|_{1500}^{\infty} - \int_{1500}^{\infty} \left(\frac{e^{-0.002y}}{-0.002}\right) dy$$

$$= \frac{1500}{0.002}e^{-3} - \left(\frac{e^{-0.002y}}{(-0.002)(-0.002)}\Big|_{1500}^{\infty}\right)$$

$$= \frac{1500}{0.002}e^{-3} + \frac{e^{-3}}{(0.002)(0.002)}$$

$$= \frac{e^{-3}}{0.002}(2000)$$

With the constant $0.002e^3$ reapplied,

$$E(Y \mid X = 1500) = 2000$$

Practical Interpretation: If the connect time is 1500 ms, then the expected time to be authorized is 2000 ms.

For the discrete random variables in Example 5.1, the conditional mean of Y given $X = 3$ is obtained from the conditional distribution in Example 5.5:

$$E(Y \mid 3) = \mu_{Y|3} = 1(0.494) + 2(0.364) + 3(0.091) + 4(0.091) = 1.86$$

The conditional mean is interpreted as the expected response time given that one bar of signal is present. The conditional variance of Y given $X = 3$ is

$$V(Y \mid 3) = (1-1.86)^2\, 0.494 + (2-1.86)^2\, 0.364 + (3-1.86)^2\, 0.091 + (4-1.86)^2\, 0.091 = 0.91$$

Independence In some random experiments, knowledge of the values of X does not change any of the probabilities associated with the values for Y. In this case, marginal probability distributions can be used to calculate probabilities more easily.

EXAMPLE 5.8 | Independent Random Variables

An orthopedic physician's practice considers the number of errors in a bill and the number of X-rays listed on the bill. There may or may not be a relationship between these random variables. Let the random variables X and Y denote the number of errors and the number of X-rays on a bill, respectively.

Assume that the joint probability distribution of X and Y is defined by $f_{XY}(x, y)$ in Figure 5.9(a). The marginal probability distributions of X and Y are also shown in Figure 5.9(a). Note that

$$f_{XY}(x, y) = f_X(x)\, f_Y(y).$$

The conditional probability mass function $f_{Y|x}(y)$ is shown in Fig. 5.9(b). Notice that for any x, $f_{Y|x}(y) = f_Y(y)$. That is, knowledge of whether or not the bill has errors does not change the probability of the number of X-rays listed on the bill.

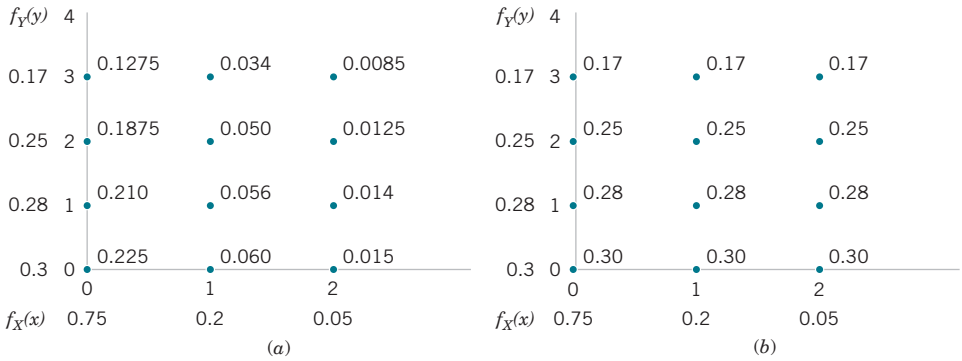


FIGURE 5.9

(a) Joint and marginal probability distributions of X and Y. (b) Conditional probability distribution of Y given $X = x$.

By analogy with independent events, we define two random variables to be **independent** whenever $f_{XY}(x, y) = f_X(x) f_Y(y)$ for all x and y. Notice that independence implies that $f_{XY}(x, y) = f_X(x) f_Y(y)$ for *all* x and y. If we find one pair of x and y in which the equality fails, then X and Y are not independent.

If two random variables are independent, then for $f_X(x) > 0$,

$$f_{Y|x}(y) = \frac{f_{XY}(x, y)}{f_X(x)} = \frac{f_X(x) f_Y(y)}{f_X(x)} = f_Y(y)$$

With similar calculations, the following equivalent statements can be shown.

Independence

For random variables X and Y, if any one of the following properties is true, the others are also true, and X and Y are **independent**.

(1) $f_{XY}(x, y) = f_X(x) f_Y(y)$ for all x and y

(2) $f_{Y|x}(y) = f_Y(y)$ for all x and y with $f_X(x) > 0$

(3) $f_{X|y}(x) = f_X(x)$ for all x and y with $f_Y(y) > 0$

(4) $P(X \in A, Y \in B) = P(X \in A)\, P(Y \in B)$ for any sets A and B in the range of X and Y, respectively. (5.8)

Rectangular Range for (X, Y) Let D denote the set of points in two-dimensional space that receive positive probability under $f_{XY}(x, y)$. If D is not rectangular, then X and Y are not independent because knowledge of X can restrict the range of values of Y that receive positive probability. If D is rectangular, independence is possible but not demonstrated. One of the conditions in Equation 5.8 must still be verified.

The variables in Example 5.2 are not independent. This can be quickly determined because the range of (X, Y) shown in Figure 5.4 is not rectangular. Consequently, knowledge of X changes the interval of values for Y with nonzero probability.

EXAMPLE 5.9 | Independent Random Variables

Suppose that Example 5.2 is modified so that the joint probability density function of X and Y is $f_{XY}(x, y) = 2 \times 10^{-6} \exp(-0.001x - 0.002y)$ for $x \geq 0$ and $y \geq 0$. Show that X and Y are independent and determine $P(X > 1000, Y < 1000)$.

Note that the range of positive probability is rectangular so that independence is possible but not yet demonstrated. The marginal probability density function of X is

$$f_X(x) = \int_0^\infty 2 \times 10^{-6} e^{-0.001x - 0.002y}\, dy$$
$$= 0.001 e^{-0.001x} \quad \text{for} \quad x > 0$$

The marginal probability density function of Y is

$$f_Y(y) = \int_0^\infty 2 \times 10^{-6} e^{-0.001x - 0.002y}\, dx$$
$$= 0.002 e^{-0.002y} \quad \text{for} \quad y > 0$$

Therefore, $f_{XY}(x, y) = f_X(x)\, f_Y(y)$ for all x and y, and X and Y are independent.

To determine the probability requested, property (4) of Equation 5.8 can be applied along with the fact that each random variable has an exponential distribution. Therefore,

$$P(X > 1000, Y < 1000) = P(X > 1000)\, P(Y < 1000)$$
$$= e^{-1}(1 - e^{-2}) = 0.318$$

Often, based on knowledge of the system under study, random variables are assumed to be independent. Then probabilities involving both variables can be determined from the marginal probability distributions. For example, the time to complete a computer search should be independent of an adult's height.

EXAMPLE 5.10 | Machined Dimensions

Let the random variables X and Y denote the lengths of two dimensions of a machined part, respectively. Assume that X and Y are independent random variables, and further assume that the distribution of X is normal with mean 10.5 millimeters and variance 0.0025 (mm^2) and that the distribution of Y is normal with mean 3.2 millimeters and variance 0.0036 (mm^2). Determine the probability that $10.4 < X < 10.6$ and $3.15 < Y < 3.25$.

Because X and Y are independent,

$$P(10.4 < X < 10.6, 3.15 < Y < 3.25)$$
$$= P(10.4 < X < 10.6) \times P(3.15 < Y < 3.25)$$
$$= P\left(\frac{10.4 - 10.5}{0.05} < Z < \frac{10.6 - 10.5}{0.05}\right) \times P\left(\frac{3.15 - 3.2}{0.06} < Z < \frac{3.25 - 3.2}{0.06}\right)$$
$$= P(-2 < Z < 2)\, P(-0.833 < Z < 0.833)$$
$$= 0.568$$

where Z denotes a standard normal random variable.

Practical Interpretation: If random variables are independent, probabilities for multiple variables are often much easier to compute.

5.3 Joint Probability Distributions for More Than Two Random Variables

More than two random variables can be defined in a random experiment. Results for multiple random variables are straightforward extensions of those for two random variables. A summary is provided here.

EXAMPLE 5.11 | Machined Dimensions

Many dimensions of a machined part are routinely measured during production. Let the random variables, X_1, X_2, X_3, and X_4 denote the lengths of four dimensions of a part. Then at least four random variables are of interest in this study.

The joint probability distribution of random variables $X_1, X_2, X_3, \ldots, X_p$ can be specified with a method to calculate the probability that $X_1, X_2, X_3, \ldots, X_p$ assume a value in any region R of p-dimensional space. For continuous random variables, a **joint probability density function** $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)$ is used to determine the probability that $(X_1, X_2, X_3, \ldots, X_p) \in R$ by the multiple integral of $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)$ over the region R.

Joint Probability Density Function

A **joint probability density function** for the continuous random variables X_1, X_2, X_3, … , X_p, denoted as $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)$, satisfies the following properties:

(1) $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p) \geq 0$

(2) $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1 dx_2 \ldots dx_p = 1$$

(3) For any region B of p-dimensional space,

$$P[(X_1, X_2, \ldots, X_p) \in B] = \int\cdots\int_B f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1 dx_2 \ldots dx_p \qquad (5.9)$$

Typically, $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)$ is defined over all of p-dimensional space by assuming that $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p) = 0$ for all points for which $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)$ is not specified.

EXAMPLE 5.12 | Component Lifetimes

In an electronic assembly, let the random variables X_1, X_2, X_3, X_4 denote the lifetime of four components, respectively, in hours. Suppose that the joint probability density function of these variables is

$$f_{X_1X_2X_3X_4}(x_1, x_2, x_3, x_4) = 9 \times 10^{-12} e^{-0.001x_1 - 0.002x_2 - 0.0015x_3 - 0.003x_4}$$
$$\text{for } x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0$$

What is the probability that the device operates for more than 1000 hours without any failures? The requested probability is $P(X_1 > 1000, X_2 > 1000, X_3 > 1000, X_4 > 1000)$, which equals the multiple integral of $f_{X_1X_2X_3X_4}(x_1, x_2, x_3, x_4)$ over the region $x_1 > 1000, x_2 > 1000, x_3 > 1000, x_4 > 1000$. The joint probability density function can be written as a product of exponential functions, and each integral is the simple integral of an exponential function. Therefore,

$$P(X_1 > 1000, X_2 > 1000, X_3 > 1000, X_4 > 1000)$$
$$= e^{-1-2-1.5-3} = 0.00055$$

Suppose that the joint probability density function of several continuous random variables is a constant c over a region R (and zero elsewhere). In this special case,

$$\iint_R \cdots \int f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)dx_1\, dx_2 \ldots dx_p = c \times (\text{volume of region } R) = 1$$

by property (2) of Equation 5.9. Therefore, $c = 1/(\text{volume of region } R)$. Furthermore, by property (3) of Equation 5.9, $P[(X_1, X_2, \ldots, X_p) \in B]$

$$= \iint_B \cdots \int f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)dx_1\, dx_2 \ldots dx_p = c \times \text{volume } (B \cap R) = \frac{\text{volume } (B \cap R)}{\text{volume } (R)}$$

When the joint probability density function is constant, the probability that the random variables assume a value in the region B is just the ratio of the volume of the region $B \cap R$ to the volume of the region R for which the probability is positive.

EXAMPLE 5.13 | Probability as a Ratio of Volumes

Suppose that the joint probability density function of the continuous random variables X and Y is constant over the region $x^2 + y^2 \le 4$. Determine the probability that $X^2 + Y^2 \le 1$.

The region that receives positive probability is a circle of radius 2. Therefore, the area of this region is 4π. The area of the region $x^2 + y^2 \le 1$ is π. Consequently, the requested probability is $\pi/(4\pi) = 1/4$.

Marginal Probability Density Function

If the joint probability density function of continuous random variables $X_1, X_2, \ldots, X_p$ is $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)$, the **marginal probability density function** of X_i is

$$f_{X_i}(x_i) = \iint \cdots \int f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1 dx_2 \ldots dx_{i-1} dx_{i+1} \ldots dx_p \qquad (5.10)$$

where the integral is over all points in the range of $X_1, X_2, \ldots, X_p$ for which $X_i = x_i$.

As for two random variables, a probability involving only one random variable, for example, $P(a < X_i < b)$, can be determined from the marginal probability distribution of X_i or from the joint probability distribution of $X_1, X_2, \ldots, X_p$. That is,

$$P(a < X_i < b) = P(-\infty < X_1 < \infty, \ldots, -\infty < X_{i-1} < \infty, a < X_i < b, -\infty < X_{i+1} < \infty, \ldots, -\infty < X_p < \infty)$$

Furthermore, $E(X_i)$ and $V(X_i)$ for $i = 1, 2, \ldots, p$ can be determined from the marginal probability distribution of X_i or from the joint probability distribution of $X_1, X_2, \ldots, X_p$ as follows.

Mean and Variance from Joint Distribution

$$E(X_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)dx_1 dx_2 \ldots dx_p = \int_{-\infty}^{\infty} x_i f_{X_i}(x_i)dx_i \qquad (5.11)$$

and

$$V(X_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_i - \mu_{X_i})^2 f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1 dx_2 \ldots dx_p$$

$$= \int_{-\infty}^{\infty} (x_i - \mu_{X_i})^2 f_{X_i}(x_i)dx_i$$

EXAMPLE 5.14 | Marginal Probability Distribution

Points that have positive probability in the joint probability distribution of three random variables X_1, X_2, X_3 are shown in Figure 5.10. Suppose the 10 points are equally likely with probability 0.1 each. The range is the nonnegative integers with $x_1 + x_2 + x_3 = 3$. The marginal probability distribution of X_2 is found as follows.

$$P(X_2 = 0) = f_{X_1X_2X_3}(3,0,0) + f_{X_1X_2X_3}(0,0,3) + f_{X_1X_2X_3}(1,0,2) + f_{X_1X_2X_3}(2,0,1) = 0.4$$

$$P(X_2 = 1) = f_{X_1X_2X_3}(2,1,0) + f_{X_1X_2X_3}(0,1,2) + f_{X_1X_2X_3}(1,1,1) = 0.3$$

$$P(X_2 = 2) = f_{X_1X_2X_3}(1,2,0) + f_{X_1X_2X_3}(0,2,1) = 0.2$$

$$P(X_2 = 3) = f_{X_1X_2X_3}(0,3,0) = 0.1$$

Also, $E(X_2) = 0(0.4) + 1(0.3) + 2(0.2) + 3(0.1) = 1$

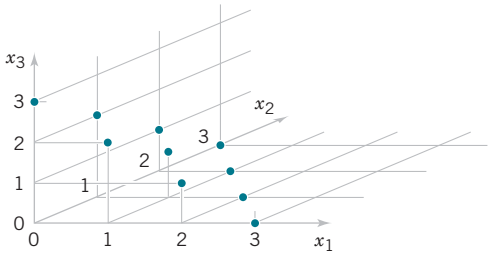


FIGURE 5.10

Joint probability distribution of X_1, X_2, and X_3. Points are equally likely.

With several random variables, we might be interested in the probability distribution of some subset of the collection of variables. The probability distribution of $X_1, X_2, \ldots, X_k$, $k < p$ can be obtained from the joint probability distribution of $X_1, X_2, \ldots, X_p$ as follows.

Distribution of a Subset of Random Variables

If the joint probability density function of continuous random variables $X_1, X_2, \ldots, X_p$ is $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)$, the probability density function of $X_1, X_2, \ldots, X_k$, $k < p$, is

$$f_{X_1X_2\ldots X_k}(x_1, x_2, \ldots, x_k) = \iint \cdots \int f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_{k+1} dx_{k+2} \ldots dx_p \qquad (5.12)$$

where the integral is over all points R in the range of $X_1, X_2, \ldots, X_p$ for which $X_1 = x_1$, $X_2 = x_2, \ldots, X_k = x_k$.

Conditional Probability Distribution Conditional probability distributions can be developed for multiple random variables by an extension of the ideas used for two random variables. For example, the joint conditional probability distribution of X_1, X_2, and X_3 given ($X_4 = x_4$, $X_5 = x_5$) is

$$f_{X_1X_2X_3 \mid x_4x_5}(x_1, x_2, x_3) = \frac{f_{X_1X_2X_3X_4X_5}(x_1, x_2, x_3, x_4, x_5)}{f_{X_4X_5}(x_4, x_5)} \quad \text{for} \quad f_{X_4X_5}(x_4, x_5) > 0.$$

The concept of independence can be extended to multiple random variables.

Independence

Random variables $X_1, X_2, \ldots, X_p$ are **independent** if and only if

$$f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, \ldots, x_p) = f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_p}(x_p) \text{ for } \textit{all } x_1, x_2, \ldots, x_p \qquad (5.13)$$

Similar to the result for only two random variables, independence implies that Equation 5.13 holds for *all* $x_1, x_2, \ldots, x_p$. If we find one point for which the equality fails, $X_1, X_2, \ldots, X_p$ are not independent. It is left as an exercise to show that if $X_1, X_2, \ldots, X_p$ are independent,

$$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_p \in A_p) = P(X_1 \in A_1)\, P(X_2 \in A_2) \ldots P(X_p \in A_p)$$

for *any* regions $A_1, A_2, \ldots, A_p$ in the range of $X_1, X_2, \ldots, X_p$, respectively.

EXAMPLE 5.15 | Negative Binomial Distribution

In Chapter 3, we showed that a negative binomial random variable with parameters p and r can be represented as a sum of r geometric random variables $X_1, X_2, \ldots, X_r$. Each geometric random variable represents the additional trials required to obtain the next success. Because the trials in a binomial experiment are independent, $X_1, X_2, \ldots, X_r$ are independent random variables.

5.4 Covariance and Correlation

When two or more random variables are defined on a probability space, it is useful to describe how they vary together; that is, it is useful to measure the relationship between the variables. A common measure of the relationship between two random variables is the **covariance**. To define the covariance, we need to describe the expected value of a function of two random variables $h(X, Y)$. The definition simply extends the one for a function of a single random variable.

Expected Value of a Function of Two Random Variables

$$E[h(X,Y)] = \begin{cases} \sum\sum h(x,y)\, f_{XY}(x,y) & X, Y \text{ discrete} \\ \int\int h(x,y)\, f_{XY}(x,y)\, dx\, dy & X, Y \text{ continuous} \end{cases} \tag{5.14}$$

That is, $E[h(X, Y)]$ can be thought of as the weighted average of $h(x, y)$ for each point in the range of (X, Y). The value of $E[h(X, Y)]$ represents the average value of $h(X, Y)$ that is expected in a long sequence of repeated trials of the random experiment.

EXAMPLE 5.16 | Expected Value of a Function of Two Random Variables

For the joint probability distribution of the two random variables in Example 5.1, calculate $E[(X - \mu_X)(Y - \mu_Y)]$.

The result is obtained by multiplying $x - \mu_X$ times $y - \mu_Y$, times $f_{XY}(x, y)$ for each point in the range of (X, Y). First, μ_X and μ_Y were determined previously from the marginal distributions for X and Y:

$$\mu_X = 2.35$$

and

$$\mu_Y = 2.49$$

Therefore,

$$\begin{aligned} E[(X - \mu_X)(Y - \mu_Y)] &= (1 - 2.35)(1 - 2.49)(0.01) \\ &+ (2 - 2.35)(1 - 2.49)(0.02) + (3 - 2.35)(1 - 2.49)(0.25) \\ &+ (1 - 2.35)(2 - 2.49)(0.02) + (2 - 2.35)(2 - 2.49)(0.03) \\ &+ (3 - 2.35)(2 - 2.49)(0.2) + (1 - 2.35)(3 - 2.49)(0.02) \\ &+ (2 - 2.35)(3 - 2.49)(0.1) + (3 - 2.35)(3 - 2.49)(0.05) \\ &+ (1 - 2.35)(4 - 2.49)(0.15) + (2 - 2.35)(4 - 2.49)(0.1) \\ &+ (3 - 2.35)(4 - 2.49)(0.05) = -0.5815 \end{aligned}$$

The covariance is defined for both continuous and discrete random variables by the same formula.

Covariance

The **covariance** between the random variables X and Y, denoted as $\text{cov}(X, Y)$ or σ_{XY}, is

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X\mu_Y \tag{5.15}$$

If the points in the joint probability distribution of X and Y that receive positive probability tend to fall along a line of positive (or negative) slope, σ_{XY}, is positive (or negative). If the points tend to fall along a line of positive slope, X tends to be greater than μ_X when Y is greater than μ_Y. Therefore, the product of the two terms $x - \mu_X$ and $y - \mu_Y$ tends to be positive. However, if the points tend to fall along a line of negative slope, $x - \mu_X$ tends to be positive when $y - \mu_Y$ is negative, and vice versa. Therefore, the product of $x - \mu_X$ and $y - \mu_Y$ tends to be negative. In this sense, the covariance between X and Y describes the variation between the two random variables. Figure 5.11 assumes all points are equally likely and shows examples of pairs of random variables with positive, negative, and zero covariance.

Covariance is a measure of *linear relationship* between the random variables. If the relationship between the random variables is nonlinear, the covariance might not be sensitive to the relationship. This is illustrated in Figure 5.11(d). The only points with nonzero probability are the points on the circle. There is an identifiable relationship between the variables. Still, the covariance is zero.

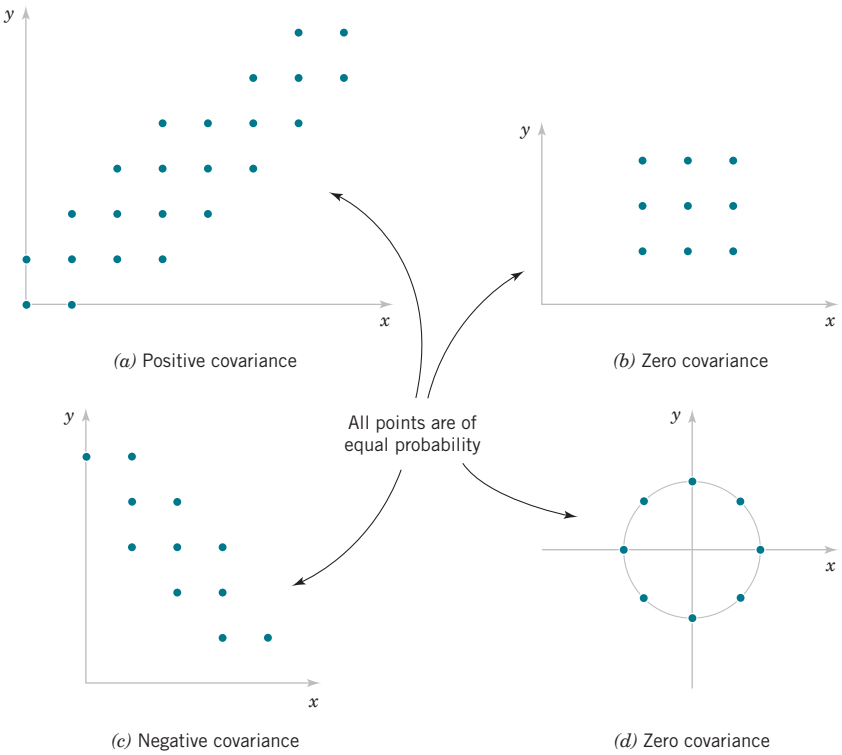


FIGURE 5.11

Joint probability distributions and the sign of covariance between X and Y.

The equality of the two expressions for covariance in Equation 5.15 is shown for continuous random variables as follows. By writing the expectations as integrals,

$$E[(Y-\mu_Y)(X-\mu_X)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f_{XY}(x,y)\,dx\,dy$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[xy-\mu_X y-x\mu_Y+\mu_X\mu_Y]f_{XY}(x,y)\,dx\,dy$$

Now

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mu_X y\,f_{XY}(x,y)\,dx\,dy = \mu_X\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf_{XY}(x,y)\,dx\,dy\right] = \mu_X\mu_Y$$

Therefore,

$$E[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf_{XY}(x,y)\,dx\,dy-\mu_X\mu_Y-\mu_X\mu_Y+\mu_X\mu_Y$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf_{XY}(x,y)\,dx\,dy-\mu_X\mu_Y = E(XY)-\mu_X\mu_Y$$

EXAMPLE 5.17

In Example 5.1, the random variables X and Y are the number of signal bars and the response time (to the nearest second), respectively. Interpret the covariance between X and Y as positive or negative.

As the signal bars increase, the response time tends to decrease. Therefore, X and Y have a negative covariance. The covariance was calculated to be −0.5815 in Example 5.16.

There is another measure of the relationship between two random variables that is often easier to interpret than the covariance.

Correlation

The **correlation** between random variables X and Y, denoted as ρ_{XY}, is

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} \tag{5.16}$$

Because $\sigma_X > 0$ and $\sigma_Y > 0$, if the covariance between X and Y is positive, negative, or zero, the correlation between X and Y is positive, negative, or zero, respectively. The following result can be shown.

For any two random variables X and Y,

$$-1 \le \rho_{XY} \le +1 \tag{5.17}$$

The correlation just scales the covariance by the product of the standard deviation of each variable. Consequently, the correlation is a dimensionless quantity that can be used to compare the linear relationships between pairs of variables in different units.

If the points in the joint probability distribution of X and Y that receive positive probability tend to fall along a line of positive (or negative) slope, ρ_{XY} is near $+1$ (or -1). If ρ_{XY} equals $+1$ or -1, it can be shown that the points in the joint probability distribution that receive positive probability fall exactly along a straight line. Two random variables with nonzero correlation are said to be **correlated**. Similar to covariance, the correlation is a measure of the linear relationship between random variables.

EXAMPLE 5.18 | Correlation

For the signal bars and response time random variables in Example 5.1, determine the correlation between the random variables X and Y.

The covariance was determined in an earlier example to equal -0.5815. From the marginal distribution of X and Y in Figure 5.6, it can be shown that $V(X) = 0.6275$ and $V(Y) = 1.4099$.

Therefore,

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{-0.5815}{\sqrt{0.6275}\sqrt{1.4099}} = -0.62$$

EXAMPLE 5.19 | Linear Relationship

Suppose that the random variable X has the following distribution: $P(X = 1) = 0.2$, $P(X = 2) = 0.6$, $P(X = 3) = 0.2$. Let $Y = 2X + 5$. That is, $P(Y = 7) = 0.2$, $P(Y = 9) = 0.6$, $P(Y = 11) = 0.2$ and so forth. Determine the correlation between X and Y.

Because X and Y are linearly related, $\rho = 1$. This can be verified by direct calculations. Try it.

For independent random variables, we do not expect any relationship in their joint probability distribution. The following result is left as an exercise.

If X and Y are independent random variables,

$$\sigma_{XY} = \rho_{XY} = 0 \tag{5.18}$$

However, if the correlation between two random variables is zero, we *cannot* conclude that the random variables are independent. Figure 5.11(d) provides an example.

5.5 Common Joint Distributions

5.5.1 Multinomial Probability Distribution

The binomial distribution can be generalized to generate a useful joint probability distribution for multiple discrete random variables. The random experiment consists of a series of independent trials. However, the outcome from each trial is categorized into one of k classes. The random variables of interest count the number of outcomes in each class.

EXAMPLE 5.20 | Digital Channel

We might be interested in a probability such as the following. Of the 20 bits received, what is the probability that 14 are excellent, 3 are good, 2 are fair, and 1 is poor? Assume that the classifications of individual bits are independent events and that the probabilities of *E, G, F*, and *P* are 0.6, 0.3, 0.08, and 0.02, respectively. One sequence of 20 bits that produces the specified numbers of bits in each class can be represented as

EEEEEEEEEEEEEEGGGFFP

Using independence, we find that the probability of this sequence is

$$P(EEEEEEEEEEEEEEGGGFFP) = 0.6^{14}0.3^{3}0.08^{2}0.02^{1}$$
$$= 2.708 \times 10^{-9}$$

Clearly, all sequences that consist of the same numbers of E's, G's, F's, and P's have the same probability. Consequently, the requested probability can be found by multiplying 2.708×10^{-9} by the number of sequences with 14 E's, 3 G's, 2 F's, and 1 P. The number of sequences is found from Chapter 2 to be

$$\frac{20!}{14!\,3!\,2!\,1!} = 2{,}325{,}600$$

Therefore, the requested probability is

$$P(14\ E\text{'s}, 3\ G\text{'s}, 2\ F\text{'s, and } 1\ P) = 2325600(2.708 \times 10^{-9})$$
$$= 0.0063$$

This example leads to the following generalization of a binomial experiment and a binomial distribution.

Multinomial Distribution

Suppose that a random experiment consists of a series of n trials. Assume that

(1) The result of each trial is classified into one of k classes.
(2) The probability of a trial generating a result in class 1, class 2, … , class k is $p_1, p_2, \ldots, p_k$, respectively, and these probabilities are constant over the trials.
(3) The trials are independent.

The random variables $X_1, X_2, \ldots, X_k$ that denote the number of trials that result in class 1, class 2, … , class k, respectively, have a **multinomial distribution** and the joint probability mass function is

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1!x_2!\ldots x_k!}\, p_1^{x_1}\, p_2^{x_2} \ldots p_k^{x_k} \qquad (5.19)$$

for $x_1 + x_2 + \cdots + x_k = n$ and $p_1 + p_2 + \cdots + p_k = 1$.

The multinomial distribution is considered a multivariable extension of the binomial distribution.

EXAMPLE 5.21 | Digital Channel

In Example 5.20, let the random variables X_1, X_2, X_3, and X_4 denote the number of bits that are E, G, F, and P, respectively, in a transmission of 20 bits. The probability that 12 of the bits received are E, 6 are G, 2 are F, and 0 are P is

$$P(X_1 = 12, X_2 = 6, X_3 = 2, X_4 = 0)$$
$$= \frac{20!}{12!\,6!\,2!\,0!}\, 0.6^{12}0.3^{6}0.08^{2}0.02^{0} = 0.0358$$

Each trial in a multinomial random experiment can be regarded as either generating or not generating a result in class i, for each $i = 1, 2, \ldots, k$. Because the random variable X_i is the number of trials that result in class i, X_i has a binomial distribution.

Mean and Variance

If $X_1, X_2, \ldots, X_k$ have a multinomial distribution, the marginal probability distribution of X_i is binomial with

$$E(X_i) = np_i \quad \text{and} \quad V(X_i) = np_i(1 - p_i) \qquad (5.20)$$

5.5.2 Bivariate Normal Distribution

An extension of a normal distribution to two random variables is an important bivariate probability distribution. The joint probability distribution can be defined to handle positive, negative, or zero correlation between the random variables.

EXAMPLE 5.22 | Bivariate Normal Distribution

At the start of this chapter, the length of different dimensions of an injection-molded part were presented as an example of two random variables. If the specifications for X and Y are 2.95 to 3.05 and 7.60 to 7.80 millimeters, respectively, we might be interested in the probability that a part satisfies both specifications; that is, $P(2.95 < X < 3.05, 7.60 < Y < 7.80)$. Each length might be modeled by a normal distribution. However, because the measurements are from the same part, the random variables are typically not independent. Therefore, a probability distribution for two normal random variables that are not independent is important in many applications.

Bivariate Normal Probability Density Function

The probability density function of a **bivariate normal distribution** is

$$f_{XY}(x, y; \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho) = \frac{1}{2\pi\sigma_X \sigma_Y \sqrt{1-\rho^2}} \times \exp\left\{ \frac{-1}{2(1-\rho^2)} \left[\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X \sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \right] \right\} \quad (5.21)$$

for $-\infty < x < \infty$ and $-\infty < y < \infty$, with parameters $\sigma_X > 0$, $\sigma_Y > 0$, $-\infty < \mu_X < \infty$, $-\infty < \mu_Y < \infty$, and $-1 < \rho < 1$.

The result that $f_{XY}(x, y; \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho)$ integrates to 1 is left as an exercise. Also, the bivariate normal probability density function is positive over the entire plane of real numbers.

Two examples of bivariate normal distributions along with corresponding *contour plots* are illustrated in Figure 5.12. Each curve on the contour plots is a set of points for which the probability density function is constant. As seen in the contour plots, the bivariate normal probability density function is constant on ellipses in the (x, y) plane. We can consider a circle to be a special case of an ellipse. The center of each ellipse is at the point (μ_X, μ_Y). If $\rho > 0$ ($\rho < 0$), the major axis of each ellipse has positive (negative) slope, respectively. If $\rho = 0$, the major axis of the ellipse is aligned with either the x or y coordinate axis.

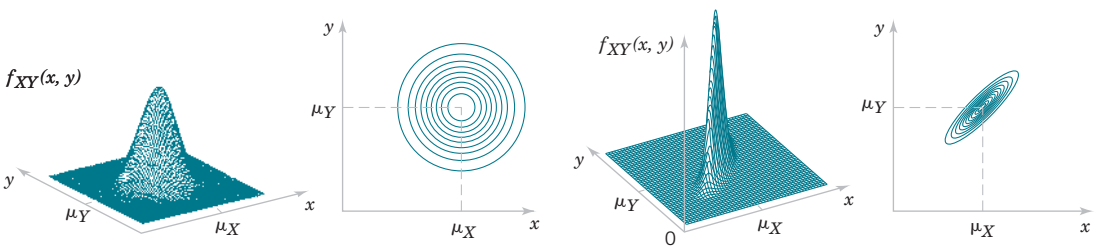


FIGURE 5.12

Examples of bivariate normal distributions.

EXAMPLE 5.23

The joint probability density function

$$f_{XY}(x, y) = \frac{1}{2\pi} e^{-0.5(x^2+y^2)}$$

is a special case of a bivariate normal distribution with $\sigma_X = 1$, $\sigma_Y = 1$, $\mu_X = 0$, $\mu_Y = 0$, and $\rho = 0$. This probability density function is illustrated in Figure 5.13. Notice that the contour plot consists of concentric circles about the origin.

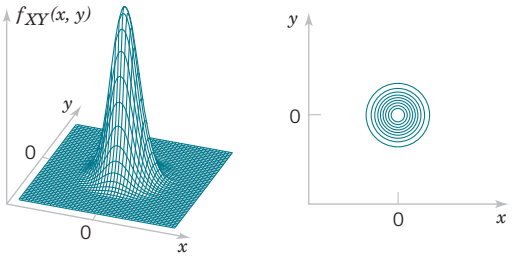


FIGURE 5.13

Bivariate normal probability density function with $\sigma_X = 1$, $\sigma_Y = 1$, $\rho = 0$, $\mu_X = 0$, and $\mu_Y = 0$.

The following results can be shown for a bivariate normal distribution. The details are left as an exercise.

Marginal Distributions of Bivariate Normal Random Variables

If X and Y have a bivariate normal distribution with joint probability density $f_{XY}(x, y, \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho)$, the marginal probability distributions of X and Y are normal with means μ_X and μ_Y and standard deviations σ_X and σ_Y, respectively. (5.22)

Conditional Distribution of Bivariate Normal Random Variables

If X and Y have a bivariate normal distribution with joint probability density $f_{XY}(x, y, \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho)$, the conditional probability distribution of Y given $X = x$ is normal with mean

$$\mu_{Y|x} = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

and variance

$$\sigma^2_{Y|x} = \sigma^2_Y (1 - \rho^2)$$

Furthermore, as the notation suggests, ρ represents the correlation between X and Y. The following result is left as an exercise.

Correlation of Bivariate Normal Random Variables

If X and Y have a bivariate normal distribution with joint probability density function $f_{XY}(x, y, \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho)$, the correlation between X and Y is ρ. (5.23)

The contour plots in Figure 5.12 illustrate that as ρ moves from 0 (left graph) to 0.9 (right graph), the ellipses narrow around the major axis. The probability is more concentrated about a line in the (xy) plane and graphically displays greater correlation between the variables. If $\rho = -1$ or $+1$, all the probability is concentrated on a line in the (x, y) plane. That is, the probability that X and Y assume a value that is not on the line is zero. In this case, the bivariate normal probability density is not defined.

In general, zero correlation does not imply independence. But in the special case that X and Y have a bivariate normal distribution, if $\rho = 0$, then X and Y are independent. The details are left as an exercise.

For Bivariate Normal Random Variables Zero Correlation Implies Independence

If X and Y have a bivariate normal distribution with $\rho = 0$, then X and Y are independent. (5.24)

An important use of the bivariate normal distribution is to calculate probabilities involving two correlated normal random variables.

EXAMPLE 5.24 | Injection-Molded Part

Suppose that the X and Y dimensions of an injection-molded part have a bivariate normal distribution with $\sigma_X = 0.04$, $\sigma_Y = 0.08$, $\mu_X = 3.00$, $\mu_Y = 7.70$, and $\rho = 0.8$. Then the probability that a part satisfies both specifications is

$$P(2.95 < X < 3.05, 7.60 < Y < 7.80)$$

This probability can be obtained by integrating $f_{XY}(x, y; \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho)$ over the region $2.95 < x < 3.05$ and $7.60 < y < 7.80$, as shown in Figure 5.3. Unfortunately, there is often no closed-form solution to probabilities involving bivariate normal distributions. In this case, the integration must be done numerically.

5.6 Linear Functions of Random Variables

A random variable is sometimes defined as a function of one or more random variables. In this section, results for linear functions are highlighted because of their importance in the remainder of the book. For example, if the random variables X_1 and X_2 denote the length and width, respectively, of a manufactured part, $Y = 2X_1 + 2X_2$ is a random variable that represents the perimeter of the part. As another example, recall that a negative binomial random variable was represented as the sum of several geometric random variables.

In this section, we develop results for random variables that are linear functions of random variables.

Linear Function

Given random variables $X_1, X_2, \ldots, X_p$ and constants $c_0, c_1, c_2, \ldots c_p$,

$$Y = c_0 + c_1X_1 + c_2X_2 + \cdots + c_pX_p \tag{5.25}$$

is a **linear function** of $X_1, X_2, \ldots, X_p$.

Now, $E(Y)$ can be found from the joint probability distribution of $X_1, X_2, \ldots, X_p$ as follows. Assume that X_1, X_2, X_p are continuous random variables. An analogous calculation can be used for discrete random variables.

$$E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(c_0 + c_1x_1 + c_2x_2 + \cdots + c_px_p) f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_p$$

$$= c_0 + c_1 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1 f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_p$$

$$+ c_2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_2 f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_p$$

$$+ , \ldots ,$$

$$+ c_p \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_p f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_p$$

By using Equation 5.11 for each of the terms in this expression, we obtain the following.

Mean of a Linear Function

If $Y = c_0 + c_1X_1 + c_2X_2 + \cdots + c_pX_p$,

$$E(Y) = c_0 + c_1E(X_1) + c_2E(X_2) + \cdots + c_pE(X_p) \tag{5.26}$$

Furthermore, it is left as an exercise to show the following.

Variance of a Linear Function

If $X_1, X_2, \ldots, X_p$ are random variables, and $Y = c_0 + c_1X_1 + c_2X_2 + \cdots + c_pX_p$, then in general,

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots + c_p^2V(X_p) + 2\sum_{i<j}\sum c_ic_j \operatorname{cov}(X_i, X_j) \tag{5.27}$$

If $X_1, X_2, \ldots, X_p$ are *independent*,

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots + c_p^2V(X_p) \tag{5.28}$$

Note that the result for the variance in Equation 5.28 requires the random variables to be independent. To see why the independence is important, consider the following simple example. Let X_1 denote any random variable and define $X_2 = -X_1$. Clearly, X_1 and X_2 are not independent. In fact, $\rho_{XY} = -1$. Now, $Y = X_1 + X_2$ is 0 with probability 1. Therefore, $V(Y) = 0$ regardless of the variances of X_1 and X_2.

EXAMPLE 5.25 | Negative Binomial Distribution

In Chapter 3, we found that if Y is a negative binomial random variable with parameters p and r, $Y = X_1 + X_2 + \cdots + X_r$, where each X_i is a geometric random variable with parameter p, and they are independent. Therefore, $E(X_i) = 1/p$ and $V(X_i) = (1-p)/p^2$. From Equation 5.26, $E(Y) = r/p$, and from Equation 5.28, $V(Y) = r(1-p)/p^2$.

An approach similar to the one applied in Example 5.25 can be used to verify the formulas for the mean and variance of an Erlang random variable in Chapter 4. An important use of Equation 5.28 is in **error propagation**, which is presented in the following example.

EXAMPLE 5.26 | Error Propagation

A semiconductor product consists of three layers. Suppose that the variances in thickness of the first, second, and third layers are 25, 40, and 30 square nanometers, respectively, and the layer thicknesses are independent. What is the variance of the thickness of the final product?

Let X_1, X_2, X_3, and X be random variables that denote the thicknesses of the respective layers and the final product. Then,

$$X = X_1 + X_2 + X_3$$

The variance of X is obtained from Equation 5.28:

$$V(X) = V(X_1) + V(X_2) + V(X_3) = 25 + 40 + 30 = 95 \text{ nm}^2$$

Consequently, the standard deviation of thickness of the final product is $95^{1/2} = 9.75$ nm, and this shows how the variation in each layer is propagated to the final product.

The particular linear function that represents the average of random variables with identical means and variances is used quite often in subsequent chapters. We highlight the results for this special case.

Mean and Variance of an Average

If $\overline{X} = (X_1 + X_2 + \cdots + X_p)/p$ with $E(X_i) = \mu \quad \text{for} \quad i = 1, 2, \ldots, p$,

$$E(\overline{X}) = \mu \tag{5.29a}$$

If $X_1, X_2, \ldots, X_p$ are also independent with $V(X_i) = \sigma^2 \quad \text{for} \quad i = 1, 2, \ldots, p$,

$$V(\overline{X}) = \frac{\sigma^2}{p} \tag{5.29b}$$

The conclusion for $V(\overline{X})$ is obtained as follows. Using Equation 5.28 with $c_i = 1/p$ and $V(X_i) = \sigma^2$ yields

$$V(\overline{X}) = \underbrace{(1/p)^2\sigma^2 + \cdots + (1/p)^2\sigma^2}_{p \text{ terms}} = \sigma^2/p$$

Another useful result concerning linear functions of random variables is a **reproductive property** that holds for independent, normal random variables.

Reproductive Property of the Normal Distribution

If $X_1, X_2, \ldots, X_p$ are independent, normal random variables with $E(X_i) = \mu_i$ and $V(X_i) = \sigma_i^2$, for $i = 1, 2, \ldots, p$, then

$$Y = c_0 + c_1X_1 + c_2X_2 + \cdots + c_pX_p$$

is a normal random variable with

$$E(Y) = c_0 + c_1\mu_1 + c_2\mu_2 + \cdots + c_p\mu_p$$

and

$$V(Y) = c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \cdots + c_p^2\sigma_p^2 \tag{5.30}$$

The mean and variance of Y follow from Equations 5.26 and 5.28. The fact that Y has a normal distribution can be obtained from moment-generating functions in the online material.

EXAMPLE 5.27 | Linear Function of Independent Normal Random Variables

Let the random variables X_1 and X_2 denote the length and width, respectively, of a manufactured part. Assume that X_1 is normal with $E(X_1) = 2$ cm and standard deviation 0.1 cm and that X_2 is normal with $E(X_2) = 5$ cm and standard deviation 0.2 cm. Also assume that X_1 and X_2 are independent. Determine the probability that the perimeter exceeds 14.5 cm.

Then $Y = 2X_1 + 2X_2$ is a normal random variable that represents the perimeter of the part. From Equation 5.30, $E(Y) = 14$ cm and

$$V(Y) = 4 \times 0.1^2 + 4 \times 0.2^2 = 0.2 \text{ cm}^2$$

Now,

$$P(Y > 14.5) = P\left(\frac{Y - \mu_Y}{\sigma_Y} > \frac{14.5 - 14}{\sqrt{0.2}}\right)$$
$$= P(Z > 1.12) = 0.13$$

5.7 General Functions of Random Variables

In many situations in statistics, it is necessary to derive the probability distribution of a function of one or more random variables. In this section, we present some results that are helpful in solving this problem.

Suppose that X is a discrete random variable with probability distribution $f_X(x)$. Let $Y = h(X)$ be a function of X that defines a one-to-one transformation between the values of X and Y and that we wish to find the probability distribution of Y. By a one-to-one transformation, we mean that each value x is related to one and only one value of $y = h(x)$ and that each value of y is related to one and only one value of x, say, $x = u(y)$ where $u(y)$ is found by solving $y = h(x)$ for x in terms of y.

Now the random variable Y takes on the value y when X takes on the value $u(y)$. Therefore, the probability distribution of Y is

$$f_Y(y) = P(Y = y) = P[X = u(y)] = f_X[u(y)]$$

We may state this result as follows.

General Function of a Discrete Random Variable

Suppose that X is a *discrete* random variable with probability distribution $f_X(x)$. Let $Y = h(X)$ define a one-to-one transformation between the values of X and Y so that the equation $y = h(x)$ can be solved uniquely for x in terms of y. Let this solution be $x = u(y)$. Then the probability mass function of the random variable Y is

$$f_Y(y) = f_X[u(y)] \tag{5.31}$$

EXAMPLE 5.28 | Function of a Discrete Random Variable

Let X be a geometric random variable with probability distribution

$$f_X(x) = p\,(1-p)^{x-1}, \qquad x = 1, 2, \ldots$$

Find the probability distribution of $Y = X^2$.

Because $X \geq 0$, the transformation is one to one; that is, $y = x^2$ and $x = \sqrt{y}$. Therefore, Equation 5.31 indicates that the distribution of the random variable Y is

$$f_Y(y) = f(\sqrt{y}) = p\,(1-p)^{\sqrt{y}-1}, \qquad y = 1, 4, 9, 16, \ldots$$

We now consider the situation in which the random variables are continuous. Let $Y = h(X)$ with X continuous and the transformation one to one.

General Function of a Continuous Random Variable

Suppose that X is a *continuous* random variable with probability distribution $f_X(x)$. The function $Y = h(X)$ is a one-to-one transformation between the values of Y and X, so that the equation $y = h(x)$ can be uniquely solved for x in terms of y. Let this solution be $x = u(y)$. The probability distribution of Y is

$$f_Y(y) = f_X[u(y)]\,|J| \tag{5.32}$$

where $J = u'(y)$ is called the **Jacobian** of the transformation and the absolute value of J is used.

Equation 5.32 is shown as follows. Let the function $y = h(x)$ be an increasing function of x. Now,

$$P(Y \le a) = P[X \le u(a)] = \int_{-\infty}^{u(a)} f_X(x)\,dx$$

If we change the variable of integration from x to y by using $x = u(y)$, we obtain $dx = u'(y)\,dy$ and then

$$P(Y \le a) = \int_{-\infty}^{a} f_X[u(y)]\,u'(y)\,dy$$

Because the integral gives the probability that $Y \le a$ for all values of a contained in the feasible set of values for y, $f_X[u(y)]\,u'(y)$ must be the probability density of Y. Therefore, the probability distribution of Y is

$$f_Y(y) = f_X[u(y)]\,u'(y) = f_X[u(y)]\,J$$

If the function $y = h(x)$ is a decreasing function of x, a similar argument holds.

EXAMPLE 5.29 | Function of a Continuous Random Variable

Let X be a continuous random variable with probability distribution

$$f_X(x) = \frac{x}{8}, \quad 0 \le x < 4$$

Find the probability distribution of $Y = h(X) = 2X + 4$.

Note that $y = h(x) = 2x + 4$ is an increasing function of x. The inverse solution is $x = u(y) = (y - 4)/2$, and from this we find the Jacobian to be $J = u'(y) = dx/dy = 1/2$. Therefore, from Equation 5.32, the probability distribution of Y is

$$f_Y(y) = \frac{(y-4)/2}{8}\left(\frac{1}{2}\right) = \frac{y-4}{32}, \qquad 4 \le y \le 12$$

5.8 Moment-Generating Functions

Suppose that X is a random variable with mean μ. Throughout this book we have used the idea of the expected value of the random variable X, and in fact $E(X) = \mu$. Now suppose that we are interested in the expected value of a function of X, $g(X) = X^r$. The expected value of this function, or $E[g(X)] = E(X^r)$, is called the rth moment about the origin of the random variable X, which we denote by μ'_r.

Definition of Moments about the Origin

The rth **moment about the origin** of the random variable X is

$$\mu'_r = E(X^r) = \begin{cases} \sum_x x^r f(x), & X \text{ discrete} \\ \int_{-\infty}^{\infty} x^r f(x)\,dx, & X \text{ continuous} \end{cases} \tag{5.33}$$

Notice that the first moment about the origin is just the mean, that is, $E(X) = \mu'_1 = \mu$. Furthermore, since the second moment about the origin is $E(X)^2 = \mu'_2$, we can write the variance of a random variable in terms of origin moments as follows:

$$\sigma^2 = E(X^2) - [E(X)]^2 = \mu'_2 - \mu^2$$

The moments of a random variable can often be determined directly from the definition in Equation 5.33, but there is an alternative procedure that is frequently useful that makes use of a special function.

Definition of a Moment-Generating Function

The **moment-generating function** of the random variable X is the expected value of e^{tX} and is denoted by $M_X(t)$. That is,

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_x e^{tx} f(x), & X \text{ discrete} \\ \int_{-\infty}^{\infty} e^{tx} f(x)\,dx, & X \text{ continuous} \end{cases} \tag{5.34}$$

The moment-generating function $M_X(t)$ will exist only if the sum or integral in the above definition converges. If the moment-generating function of a random variable does exist, it can be used to obtain all the origin moments of the random variable.

Let X be a random variable with moment-generating function $M_X(t)$. Then

$$\mu'_r = \left.\frac{d^r M_X(t)}{dt_r}\right|_{t=0} \tag{5.35}$$

Assuming that we can differentiate inside the summation and integral signs,

$$\frac{d^r M_X(t)}{dt^r} = \begin{cases} \sum_x x^r e^{tx} f(x), & X \text{ discrete} \\ \int_{-\infty}^{\infty} x^r e^{tx} f(x)\,dx, & X \text{ continuous} \end{cases}$$

Now if we set $t = 0$ in this expression, we find that

$$\left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0} = E(X^r)$$

EXAMPLE 5.30 | Moment-Generating Function for a Binomial Random Variable

Suppose that X has a binomial distribution, that is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0,1,\ldots,n$$

Determine the moment-generating function and use it to verify that the mean and variance of the binomial random variable are $\mu = np$ and $\sigma^2 = np(1-p)$.

From the definition of a moment-generating function, we have

$$M_X(t) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x}$$

This last summation is the binomial expansion of $[pe^t + (1-p)]^n$, so

$$M_X(t) = [pe^t + (1-p)]^n$$

Taking the first and second derivatives, we obtain

$$M'_X(t) = \frac{dM_X(t)}{dt} = npe^t[1 + p(e^t - 1)]^{n-1}$$

and

$$M''_X(t) = \frac{d^2M_X(t)}{dt^2} = npe^t(1 - p + npe^t)[1 + p(e^t - 1)]^{n-2}$$

If we set $t = 0$ in $M'_X(t)$, we obtain

$$M'_X(t)|_{t=0} = \mu'_1 = \mu = np$$

which is the mean of the binomial random variable X. Now if we set $t = 0$ in $M''_X(t)$,

$$M''_X(t)|_{t=0} = \mu'_2 = np(1 - p + np)$$

Therefore, the variance of the binomial random variable is

$$\begin{aligned}\sigma^2 &= \mu'_2 - \mu^2 = np(1 - p + np) - (np)^2 \\ &= np - np^2 = np(1-p)\end{aligned}$$

EXAMPLE 5.31 | Moment-Generating Function for a Normal Random Variable

Find the moment-generating function of the normal random variable and use it to show that the mean and variance of this random variable are μ and σ^2, respectively.

The moment-generating function is

$$\begin{aligned}M_X(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-[x^2 - 2(\mu + t\sigma^2)x + \mu^2]/(2\sigma^2)} dx\end{aligned}$$

If we complete the square in the exponent, we have

$$x^2 - 2(\mu + t\sigma^2)x + \mu^2 = [x - (\mu + t\sigma^2)]^2 - 2\mu t\sigma^2 - t^2\sigma^4$$

and then

$$\begin{aligned}M_X(t) &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\{[x-(\mu+t\sigma^2)]^2 - 2\mu t\sigma^2 - t^2\sigma^4\}/(2\sigma^2)} dx \\ &= e^{\mu t + \sigma^2 t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)[x-(\mu+t\sigma^2)]^2/\sigma^2} dx\end{aligned}$$

Let $u = [x - (\mu + t\sigma^2)]/\sigma$. Then $dx = \sigma du$, and the last expression above becomes

$$M_X(t) = e^{\mu t + \sigma^2 t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

Now, the integral is just the total area under a standard normal density, which is 1, so the moment-generating function of a normal random variable is

$$M_X(t) = e^{\mu t + \sigma^2 t^2/2}$$

Differentiating this function twice with respect to t and setting $t = 0$ in the result yields

$$\left.\frac{dM_X(t)}{dt}\right|_{t=0} = \mu'_1 = \mu \quad \text{and} \quad \left.\frac{d^2M_X(t)}{dt^2}\right|_{t=0} = \mu'_2 = \sigma^2 + \mu^2$$

Therefore, the variance of the normal random variable is

$$\sigma^2 = \mu'_2 - \mu^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

Moment-generating functions have many important and useful properties. One of the most important of these is the *uniqueness property*. That is, the moment-generating function of a random variable is unique when it exists, so if we have two random variables X and Y, say, with moment-generating functions $M_X(t)$ and $M_Y(t)$, then if $M_X(t) = M_Y(t)$ for all values of t,

both X and Y have the same probability distribution. Some of the other useful properties of the moment-generating function are summarized as follows.

Properties of Moment-Generating Functions

If X is a random variable and a is a constant, then

(1) $M_{X+a}(t) = e^{at}M_X(t)$

(2) $M_{aX}(t) = M_X(at)$

If $X_1, X_2, \ldots, X_n$ are independent random variables with moment-generating functions $M_{X_1}(t), M_{X_2}(t), \ldots, M_{X_n}(t)$, respectively, and if $Y = X_1 + X_2 + \cdots + X_n$, then the moment-generating function of Y is

(3) $M_Y(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdot \ldots \cdot M_{X_n}(t)$ (5.36)

Property (1) follows from $M_{X+a}(t) = E[e^{t(X+a)}] = e^{at}E(e^{tX}) = e^{at}M_X(t)$. Property (2) follows from $M_{aX}(t) = E[e^{t(aX)}] = E[e^{(at)X}] = M_X(at)$. Consider property (3) for the case where the X's are continuous random variables:

$$M_Y(t) = E(e^{tY}) = E[e^{t(X_1+X_2+\cdots+X_n)}]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{t(x_1+x_2+\cdots+x_n)}f(x_1, x_2, \ldots, x_n)\,dx_1\,dx_2\ldots dx_n$$

Because the X's are independent,

$$f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot \cdots \cdot f_{X_n}(x_n)$$

and one may write

$$M_Y(t) = \int_{-\infty}^{\infty} e^{tx_1}f_{X_1}(x_1)\,dx_1 \int_{-\infty}^{\infty} e^{tx_2}f_{X_2}(x_2)\,dx_2 \cdots \int_{-\infty}^{\infty} e^{tx_n}f_{X_n}(x_n)\,dx_n$$
$$= M_{X_1}(t) \cdot M_{X_2}(t) \cdot \cdots \cdot M_{X_n}(t)$$

For the case when the X's are discrete, we would use the same approach replacing integrals with summations.

Equation 5.36 is particularly useful. In many situations we need to find the distribution of the sum of two or more independent random variables, and often this result makes the problem very easy. This is illustrated in the following example.

EXAMPLE 5.32 | Distribution of a Sum of Poisson Random Variables

Suppose that X_1 and X_2 are two independent Poisson random variables with parameters λ_1 and λ_2, respectively. Determine the probability distribution of $Y = X_1 + X_2$.

The moment-generating function of a Poisson random variable with parameter λ is

$$M_X(t) = e^{\lambda(e^t-1)}$$

so the moment-generating functions of X_1 and X_2 are $M_{X_1}(t) = e^{\lambda_1(e^t-1)}$ and $M_{X_2}(t) = e^{\lambda_2(e^t-1)}$, respectively. Using Equation 5.36, the moment-generating function of $Y = X_1 + X_2$ is

$$M_Y(t) = M_{X_1}(t)M_{X_2}(t) = e^{\lambda_1(e^t-1)}e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}$$

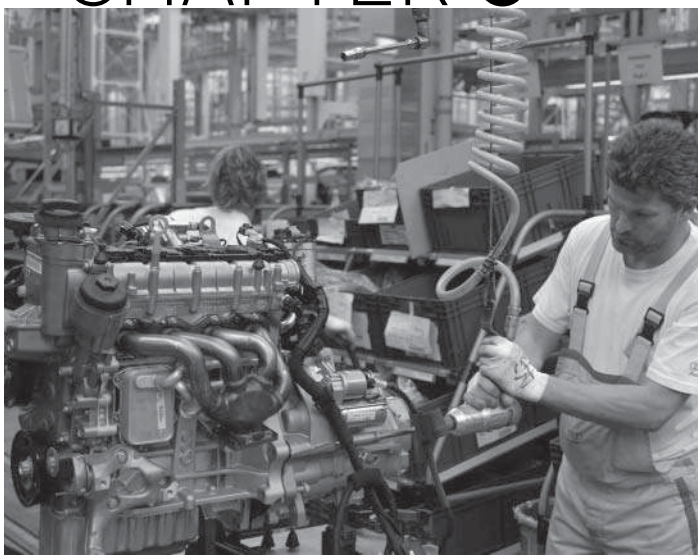
which is recognized as the moment-generating function of a Poisson random variable with parameter $\lambda_1 + \lambda_2$. Therefore, the sum of two independent Poisson random variables with parameters λ_1 and λ_2 is a Poisson random variable with parameter equal to the sum $\lambda_1 + \lambda_2$.

Important Terms and Concepts

Bivariate distribution
Bivariate normal distribution
Conditional mean
Conditional probability density function
Conditional probability mass function
Conditional variance
Correlation
Covariance
Error propagation
General functions of a random variable
Independence
Joint probability density function
Joint probability distribution
Joint probability mass function
Linear functions of random variables
Marginal probability distribution
Moment-generating function
Multinomial distribution
Reproductive property of the normal distribution

CHAPTER 6

Descriptive Statistics



Agencja Fotograficzna Caro/Alamy Stock Photo

LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Compute and interpret the sample mean, sample variance, sample standard deviation, sample median, and sample range
2. Explain the concepts of sample mean, sample variance, population mean, and population variance
3. Construct and interpret visual data displays, including the stem-and-leaf display, the histogram, and the box plot
4. Explain the concept of random sampling
5. Construct and interpret normal probability plots
6. Explain how to use box plots and other data displays to visually compare two or more samples of data
7. Know how to use simple time series plots to visually display the important features of time-oriented data
8. Know how to construct and interpret scatter diagrams of two or more variables

CHAPTER OUTLINE

6.1 Numerical Summaries of Data

6.2 Stem-and-Leaf Diagrams

6.3 Frequency Distributions and Histograms

6.4 Box Plots

6.5 Time Sequence Plots

6.6 Scatter Diagrams

6.7 Probability Plots

Statistics is the science of data. An important aspect of dealing with data is organizing and summarizing the data in ways that facilitate its interpretation and subsequent analysis. This aspect of statistics is called **descriptive statistics**, and is the subject of this chapter. For example, in Chapter 1 we presented eight prototype units made on the pull-off force of prototype automobile engine connectors. The observations (in pounds) were 12.6, 12.9, 13.4, 12.3, 13.6, 13.5, 12.6, and 13.1. There is obvious variability in the pull-off force values. How should we summarize the information in these data? This is the general question that we consider. Data summary methods should highlight the important features of the data, such as the middle or central tendency and the variability, because these characteristics are most often important for engineering decision making. We will see that there are both numerical methods for summarizing data and a number of powerful graphical techniques. The graphical techniques are particularly important. Any good statistical analysis of data should always begin with **plotting the data**.

6.1 Numerical Summaries of Data

Well-constructed data summaries and displays are essential to good statistical thinking because they can focus the engineer on important features of the data or provide insight about the type of model that should be used in solving the problem. The computer has become an important tool in the presentation and analysis of data. Although many statistical techniques require only a handheld calculator, this approach may require much time and effort, and a computer will perform the tasks much more efficiently.

Most statistical analysis is done using a prewritten library of statistical programs. The user enters the data and then selects the types of analysis and output displays that are of interest. Statistical software packages are available for both mainframe machines and personal computers. We present examples of typical output from computer software throughout the book. We do not discuss the hands-on use of specific software packages for entering and editing data or using commands.

We often find it useful to describe data features **numerically**. For example, we can characterize the location or central tendency in the data by the ordinary arithmetic average or mean. Because we almost always think of our data as a sample, we refer to the arithmetic mean as the **sample mean**.

Sample Mean

If the n observations in a sample are denoted by $x_1, x_2, \ldots, x_n$, the **sample mean** is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{6.1}$$

EXAMPLE 6.1 | Sample Mean

Let's consider the eight observations on pull-off force collected from the prototype engine connectors from Chapter 1. The eight observations are $x_1 = 12.6$, $x_2 = 12.9$, $x_3 = 13.4$, $x_4 = 12.3$, $x_5 = 13.6$, $x_6 = 13.5$, $x_7 = 12.6$, and $x_8 = 13.1$. The sample mean is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{8} x_i}{8}$$

$$= \frac{12.6 + 12.9 + \cdots + 13.1}{8} = \frac{104}{8} = 13.0 \text{ pounds}$$

A physical interpretation of the sample mean as a measure of location is shown in the **dot diagram** of the pull-off force data. See Figure 6.1. Notice that the sample mean $\bar{x} = 13.0$ can be thought of as a "balance point." That is, if each observation represents 1 pound of mass placed at the point on the x-axis, a fulcrum located at $\bar{x}$ would balance this system of weights exactly.

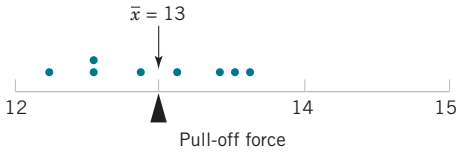


FIGURE 6.1

Dot diagram showing the sample mean as a balance point for a system of weights.

The sample mean is the average value of all observations in the data set. Usually, these data are a **sample** of observations that have been selected from some larger **population** of observations. Here the population might consist of all the connectors that will be manufactured and sold to customers. Recall that this type of population is called a **conceptual** or **hypothetical population** because it does not physically exist. Sometimes there is an actual physical population, such as a lot of silicon wafers produced in a semiconductor factory.

In previous chapters, we have introduced the mean of a probability distribution, denoted μ. If we think of a probability distribution as a model for the population, one way to think of the mean is as the average of all the measurements in the population. For a finite population with N equally likely values, the probability mass function is $f(x_i) = 1/N$ and the mean is

$$\mu = \sum_{i-1}^{N} x_i f(x_i) = \frac{\sum_{i=1}^{N} x_i}{N} \tag{6.2}$$

The sample mean, $\bar{x}$, is a reasonable estimate of the population mean, μ. Therefore, the engineer designing the connector using a 3/32-inch wall thickness would conclude on the basis of the data that an estimate of the mean pull-off force is 13.0 pounds.

Although the sample mean is useful, it does not convey all of the information about a sample of data. The variability or scatter in the data may be described by the **sample variance** or the **sample standard deviation**.

Sample Variance and Standard Deviation

If $x_1, x_2, \ldots, x_n$ is a sample of n observations, the **sample variance** is

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \tag{6.3}$$

The **sample standard deviation**, s, is the positive square root of the sample variance.

The units of measurement for the sample variance are the square of the original units of the variable. Thus, if x is measured in pounds, the units for the sample variance are $(\text{pounds})^2$. The standard deviation has the desirable property of measuring variability in the original units of the variable of interest, x.

How Does the Sample Variance Measure Variability? To see how the sample variance measures dispersion or variability, refer to Figure 6.2, which shows a dot diagram with the

deviations $x_i - \bar{x}$ for the connector pull-off force data. The higher the amount of variability in the pull-off force data, the larger in absolute magnitude some of the deviations $x_i - \bar{x}$ will be. Because the deviations $x_i - \bar{x}$ always sum to zero, we must use a measure of variability that changes the negative deviations to non-negative quantities. Squaring the deviations is the approach used in the sample variance. Consequently, if s^2 is small, there is relatively little variability in the data, but if s^2 is large, the variability is relatively large.

EXAMPLE 6.2 | Sample Variance

Table 6.1 displays the quantities needed for calculating the sample variance and sample standard deviation for the pull-off force data. These data are plotted in Figure 6.2. The numerator of s^2 is

$$\sum_{i=1}^{8} (x_i - \bar{x})^2 = 1.60$$

so the sample variance is

$$s^2 = \frac{1.60}{8-1} = \frac{1.60}{7} = 0.2286 \text{ (pounds)}^2$$

and the sample standard deviation is

$$s = \sqrt{0.2286} = 0.48 \text{ pounds}$$

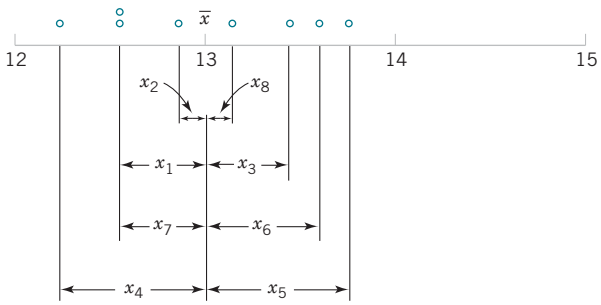


FIGURE 6.2

How the sample variance measures variability through the deviations $x_i - \bar{x}$.

TABLE 6.1 **Calculation of Terms for the Sample Variance and Sample Standard Deviation**

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	12.6	−0.4	0.16
2	12.9	−0.1	0.01
3	13.4	0.4	0.16
4	12.3	−0.7	0.49
5	13.6	0.6	0.36
6	13.5	0.5	0.25
7	12.6	−0.4	0.16
8	13.1	0.1	0.01
Total	104.0	0.0	1.60

Computation of s^2 The computation of s^2 requires calculation of $\bar{x}$, n subtractions, and n squaring and adding operations. If the original observations or the deviations $x_i - \bar{x}$ are not integers, the deviations $x_i - \bar{x}$ may be tedious to work with, and several decimals may have to be carried to ensure numerical accuracy. A more efficient computational formula for the sample variance is obtained as follows:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} \left(x_i^2 + \bar{x}^2 - 2\bar{x}x_i\right)}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 + n\bar{x}^2 - 2\bar{x}\sum_{i=1}^{n} x_i}{n-1}$$

and because $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$, this last equation reduces to

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} \tag{6.4}$$

Note that Equation 6.4 requires squaring each individual x_i, then squaring the sum of x_i, subtracting $\left(\sum x_i\right)^2/n$ from $\sum x_i^2$, and finally dividing by $n-1$. Sometimes this is called the *shortcut method for calculating* s^2 (or s).

EXAMPLE 6.3 | Shortcut Calculation for s^2

We calculate the sample variance and standard deviation using the shortcut method in Equation 6.4. The formula gives

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{1353.6 - \frac{(104)^2}{8}}{7}$$
$$= \frac{1.60}{7} = 0.2286 \text{ (pounds)}^2$$

and

$$s = \sqrt{0.2286} = 0.48 \text{ pounds}$$

These results agree exactly with those obtained previously.

Analogous to the sample variance s^2, the variability in the population is defined by the **population variance** (σ^2). As in earlier chapters, the positive square root of σ^2, or σ, denotes the **population standard deviation**. When the population is finite and consists of N equally likely values, we may define the population variance as

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N} \tag{6.5}$$

We observed previously that the sample mean could be used as an estimate of the population mean. Similarly, the sample variance is an estimate of the population variance. In Chapter 7, we discuss **estimation of parameters** more formally.

Note that the divisor for the sample variance is the sample size minus 1 ($n-1$), and for the population variance, it is the population size N. If we knew the true value of the **population mean** μ, we could find the *sample* variance as the average square deviation of the sample observations about μ. In practice, the value of μ is almost never known, and so the sum of the squared deviations about the sample average $\bar{x}$ must be used instead. However, the observations x_i tend to be closer to their average, $\bar{x}$, than to the population mean, μ. Therefore, to compensate for this, we use $n-1$ as the divisor rather than n. If we used n as the divisor in the sample variance, we would obtain a measure of variability that is on the average consistently smaller than the true population variance σ^2.

Another way to think about this is to consider the sample variance s^2 as being based on $n-1$ **degrees of freedom**. The term *degrees of freedom* results from the fact that the n deviations $x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x}$ always sum to zero, and so specifying the values of any $n-1$ of these quantities automatically determines the remaining one. This was illustrated in Table 6.1. Thus, only $n-1$ of the n deviations, $x_i - \bar{x}$, are freely determined. We may think of the number of degrees of freedom as the number of independent pieces of information in the data.

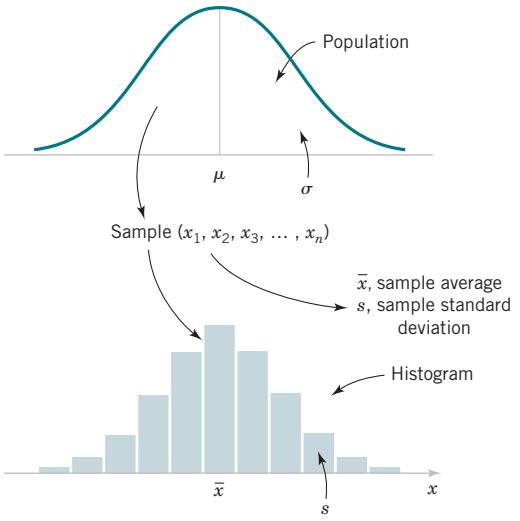


FIGURE 6.3
Relationship between a population and a sample.

In addition to the sample variance and sample standard deviation, the **sample range**, or the difference between the largest and smallest observations, is often a useful measure of variability. The sample range is defined as follows.

Sample Range

If the n observations in a sample are denoted by $x_1, x_2, \ldots, x_n$, the **sample range** is

$$r = \max(x_i) - \min(x_i) \tag{6.6}$$

For the pull-off force data, the sample range is $r = 13.6 - 12.3 = 1.3$. Generally, as the variability in sample data increases, the sample range increases.

The sample range is easy to calculate, but it ignores all of the information in the sample data between the largest and smallest values. For example, the two samples 1, 3, 5, 8, and 9 and 1, 5, 5, 5, and 9 both have the same range ($r = 8$). However, the standard deviation of the first sample is $s_1 = 3.35$, while the standard deviation of the second sample is $s_2 = 2.83$. The variability is actually less in the second sample.

Sometimes when the sample size is small, say $n < 8$ or 10, the information loss associated with the range is not too serious. For example, the range is used widely in statistical quality control where sample sizes of 4 or 5 are fairly common. We discuss some of these applications in Chapter 15.

In most statistics problems, we work with a sample of observations selected from the population that we are interested in studying. Figure 6.3 illustrates the relationship between the population and the sample.

6.2 Stem-and-Leaf Diagrams

The dot diagram is a useful data display for small samples up to about 20 observations. However, when the number of observations is moderately large, other graphical displays may be more useful.

For example, consider the data in Table 6.2. These data are the compressive strengths in pounds per square inch (psi) of 80 specimens of a new aluminum-lithium alloy undergoing evaluation as a possible material for aircraft structural elements. The data were recorded in the order of testing, and in this format they do not convey much information about compressive strength.

TABLE 6.2 Compressive Strength (in psi) of 80 Aluminum-Lithium Alloy Specimens

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

Questions such as "What percent of the specimens fail below 120 psi?" are not easy to answer. Because there are many observations, constructing a dot diagram of these data would be relatively inefficient; more effective displays are available for large data sets.

A **stem-and-leaf diagram** is a good way to obtain an informative visual display of a data set $x_1, x_2, \ldots, x_n$ where each number x_i consists of at least two digits. To construct a stem-and-leaf diagram, use the following steps.

Steps to Construct a Stem-and-Leaf Diagram

(1) Divide each number x_i into two parts: a **stem**, consisting of one or more of the leading digits, and a **leaf**, consisting of the remaining digit.
(2) List the stem values in a vertical column.
(3) Record the leaf for each observation beside its stem.
(4) Write the units for stems and leaves on the display.

To illustrate, if the data consist of percent defective information between 0 and 100 on lots of semiconductor wafers, we can divide the value 76 into the stem 7 and the leaf 6. In general, we should choose relatively few stems in comparison with the number of observations. It is usually best to choose between 5 and 20 stems.

EXAMPLE 6.4 | Alloy Strength

To illustrate the construction of a stem-and-leaf diagram, consider the alloy compressive strength data in Table 6.2. We select as stem values the numbers 7, 8, 9, … , 24. The resulting stem-and-leaf diagram is presented in Figure 6.4. The last column in the diagram is a frequency count of the number of leaves associated with each stem. Inspection of this display immediately reveals that most of the compressive strengths lie between 110 and 200 psi and that a central value is somewhere between 150 and 160 psi. Furthermore, the strengths are distributed approximately symmetrically about the central value. The stem-and-leaf diagram enables us to determine quickly some important features of the data that were not immediately obvious in the original display in Table 6.2.

In some data sets, providing more classes or stems may be desirable. One way to do this would be to modify the original stems as follows: Divide stem 5 into two new stems, 5L and 5U.

Stem 5L has leaves 0, 1, 2, 3, and 4, and stem 5U has leaves 5, 6, 7, 8, and 9. This will double the number of original stems. We could increase the number of original stems by four by defining five new stems: 5z with leaves 0 and 1, 5t (for twos and three) with leaves 2 and 3, 5f (for fours and fives) with leaves 4 and 5, 5s (for six and seven) with leaves 6 and 7, and 5e with leaves 8 and 9.

Figure 6.5 is a typical computer-generated stem-and-leaf display of the compressive strength data in Table 6.2. The software uses the same stems as in Figure 6.4. Note also that the computer orders the leaves from smallest to largest on each stem. This form of the plot is usually called an **ordered stem-and-leaf diagram**. This is not usually used when the plot is constructed manually because it can be time-consuming. The computer also adds a column to the left of the stems that provides a count of the observations at and above each stem in the upper half of the display and a count of the observations at and below each stem in the lower half of the display. At the middle stem of 16, the column indicates the number of observations at this stem.

The ordered stem-and-leaf display makes it relatively easy to find data features such as percentiles, quartiles, and the median. The **sample median** is a measure of central tendency that divides the data into two equal parts, half below the median and half above. If the number of observations is even, the median is halfway between the two central values. From Figure 6.5 we find the 40th and 41st values of strength as 160 and 163, so the median is $(160 + 163)/2 = 161.5$. If the number of observations is odd, the median is the central value. The **sample mode** is the most frequently occurring data value. Figure 6.5 indicates that the mode is 158; this value occurs four times, and no other value occurs as frequently in the sample. If there were more than one value that occurred four times, the data would have multiple modes.

Stem	Leaf	Frequency
7	6	1
8	7	1
9	7	1
10	5 1	2
11	5 8 0	3
12	1 0 3	3
13	4 1 3 5 3 5	6
14	2 9 5 8 3 1 6 9	8
15	4 7 1 3 4 0 8 8 6 8 0 8	12
16	3 0 7 3 0 5 0 8 7 9	10
17	8 5 4 4 1 6 2 1 0 6	10
18	0 3 6 1 4 1 0	7
19	9 6 0 9 3 4	6
20	7 1 0 8	4
21	8	1
22	1 8 9	3
23	7	1
24	5	1

Stem: Tens and hundreds digits (psi); Leaf: Ones digits (psi).

FIGURE 6.4

Stem-and-leaf diagram for the compressive strength data in Table 6.2.

Stem-and-leaf of Strength

$N = 80$	Stem	Unit = 1.0
1	7	6
2	8	7
3	9	7
5	10	1 5
8	11	0 5 8
11	12	0 1 3
17	13	1 3 3 4 5 5
25	14	1 2 3 5 6 8 9 9
37	15	0 0 1 3 4 4 6 7 8 8 8 8
(10)	16	0 0 0 3 3 5 7 7 8 9
33	17	0 1 1 2 4 4 5 6 6 8
23	18	0 0 1 1 3 4 6
16	19	0 3 4 6 9 9
10	20	0 1 7 8
6	21	8
5	22	1 8 9
2	23	7
1	24	5

FIGURE 6.5

A typical computer-generated stem-and-leaf diagram.

EXAMPLE 6.5 | Chemical Yield

Figure 6.6 is the stem-and-leaf diagram for 25 observations on batch yields from a chemical process. In Figure 6.6(a), we have used 6, 7, 8, and 9 as the stems. This results in too few stems, and the stem-and-leaf diagram does not provide much information about the data. In Figure 6.6(b), we have divided each stem into two parts, resulting in a display that more adequately displays the data. Figure 6.6(c) illustrates a stem-and-leaf display with each stem divided into five parts. There are too many stems in this plot, resulting in a display that does not tell us much about the shape of the data.

Stem	Leaf
6	1 3 4 5 5 6
7	0 1 1 3 5 7 8 8 9
8	1 3 4 4 7 8 8
9	2 3 5

(*a*)

Stem	Leaf
6L	1 3 4
6U	5 5 6
7L	0 1 1 3
7U	5 7 8 8 9
8L	1 3 4 4
8U	7 8 8
9L	2 3
9U	5

(*b*)

Stem	Leaf
6z	1
6t	3
6f	4 5 5
6s	6
6e	
7z	0 1 1
7t	3
7f	5
7s	7
7e	8 8 9
8z	1
8t	3
8f	4 4
8s	7
8e	8 8
9z	
9t	2 3
9f	5
9s	
9e	

(*c*)

FIGURE 6.6

Stem-and-leaf displays for Example 6.5. Stem: Tens digits; leaf: ones digits.

We can also divide data into more than two parts. When an ordered set of data is divided into four equal parts, the division points are called **quartiles**. The *first* or *lower quartile*, q_1, is a value that has approximately 25% of the observations below it and approximately 75% of the observations above. The *second quartile*, q_2, has approximately 50% of the observations below its value. The second quartile is exactly equal to the median. The *third* or *upper quartile*, q_3, has approximately 75% of the observations below its value. As in the case of the median, the quartiles may not be unique. The compressive strength data in Figure 6.5 contain $n = 80$ observations. Therefore, calculate the first and third quartiles as the $(n + 1)/4$ and $3(n + 1)/4$ ordered observations and interpolate as needed, for example, $(80 + 1)/4 = 20.25$ and $3(80 + 1)/4 = 60.75$. Therefore, interpolating between the 20th and 21st ordered observation we obtain $q_1 = 143.50$ and between the 60th and 61st observation we obtain $q_3 = 181.00$. In general, the $100k$th **percentile** is a data value such that approximately $100k\%$ of the observations are at or below this value and approximately $100(1 - k)\%$ of them are above it. Finally, we may use the

TABLE 6.3 Summary Statistics for the Compressive Strength Data from Software

N	Mean	Median	StDev	SE Mean	Min	Max	Q1	Q3
80	162.66	161.50	33.77	3.78	76.00	245.00	143.50	181.00

interquartile range, defined as IQR $= q_3 - q_1$, as a measure of variability. The interquartile range is less sensitive to the extreme values in the sample than is the ordinary sample range.

Many statistics software packages provide data summaries that include these quantities. Typical computer output for the compressive strength data in Table 6.2 is shown in Table 6.3.

6.3 Frequency Distributions and Histograms

A **frequency distribution** is a more compact summary of data than a stem-and-leaf diagram. To construct a frequency distribution, we must divide the range of the data into intervals, which are usually called **class intervals**, **cells**, or **bins**. If possible, the bins should be of equal width in order to enhance the visual information in the frequency distribution. Some judgment must be used in selecting the number of bins so that a reasonable display can be developed. The number of bins depends on the number of observations and the amount of scatter or dispersion in the data. A frequency distribution that uses either too few or too many bins will not be informative. We usually find that between 5 and 20 bins is satisfactory in most cases and that the number of bins should increase with n. Several sets of rules can be used to determine the number of bins in a histogram. However, choosing the **number of bins** approximately equal to the square root of the number of observations often works well in practice.

A frequency distribution for the comprehensive strength data in Table 6.2 is shown in Table 6.4. Because the data set contains 80 observations, and because $\sqrt{80} \simeq 9$, we suspect that about eight to nine bins will provide a satisfactory frequency distribution. The largest and smallest data values are 245 and 76, respectively, so the bins must cover a range of at least $245 - 76 = 169$ units on the psi scale. If we want the lower limit for the first bin to begin slightly below the smallest data value and the upper limit for the last bin to be slightly above the largest data value, we might start the frequency distribution at 70 and end it at 250. This is an interval or range of 180 psi units. Nine bins, each of width 20 psi, give a reasonable frequency distribution, so the frequency distribution in Table 6.4 is based on nine bins.

The second row of Table 6.4 contains a **relative frequency distribution**. The relative frequencies are found by dividing the observed frequency in each bin by the total number of observations. The last row of Table 6.4 expresses the relative frequencies on a cumulative basis.

TABLE 6.4 Frequency Distribution for the Compressive Strength Data in Table 6.2

Class	$70 \le x < 90$	$90 \le x < 110$	$110 \le x < 130$	$130 \le x < 150$	$150 \le x < 170$
Frequency	2	3	6	14	22
Relative frequency	0.0250	0.0375	0.0750	0.1750	0.2750
Cumulative relative frequency	0.0250	0.0625	0.1375	0.3125	0.5875
	$170 \le x < 190$	$190 \le x < 210$	$210 \le x < 230$	$230 \le x < 250$	
Frequency	17	10	4	2	
Relative frequency	0.2125	0.1250	0.0500	0.0250	
Cumulative relative frequency	0.8000	0.9250	0.9750	1.0000	

Frequency distributions are often easier to interpret than tables of data. For example, from Table 6.4, it is very easy to see that most of the specimens have compressive strengths between 130 and 190 psi and that 97.5 percent of the specimens fall below 230 psi.

The **histogram** is a visual display of the frequency distribution. The steps for constructing a histogram follow.

Constructing a Histogram (Equal Bin Widths)

(1) Label the bin (class interval) boundaries on a horizontal scale.
(2) Mark and label the vertical scale with the frequencies or the relative frequencies.
(3) Above each bin, draw a rectangle where height is equal to the frequency (or relative frequency) corresponding to that bin.

Figure 6.7 is the histogram for the compression strength data. The histogram, like the **stem-and-leaf diagram**, provides a visual impression of the shape of the distribution of the measurements and information about the central tendency and scatter or dispersion in the data. Notice the symmetric, bell-shaped distribution of the strength measurements in Figure 6.7. This display often gives insight about possible choices of probability distributions to use as a model for the population. For example, here we would likely conclude that the normal distribution is a reasonable model for the population of compression strength measurements.

Sometimes a histogram with **unequal bin widths** will be employed. For example, if the data have several extreme observations or outliers, using a few equal-width bins will result in nearly all observations falling in just a few of the bins. Using many equal-width bins will result in many bins with zero frequency. A better choice is to use shorter intervals in the region where most of the data fall and a few wide intervals near the extreme observations. When the bins are of unequal width, the rectangle's **area** (not its height) should be proportional to the bin frequency. This implies that the rectangle height should be

$$\text{Rectangle height} = \frac{\text{Bin frequency}}{\text{Bin width}}$$

In passing from either the original data or stem-and-leaf diagram to a frequency distribution or histogram, we have lost some information because we no longer have the individual observations. However, this information loss is often small compared with the conciseness and ease of interpretation gained in using the **frequency distribution and histogram**.

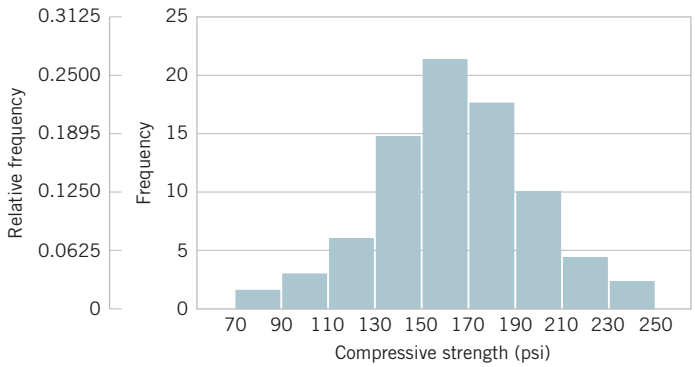


FIGURE 6.7

Histogram of compressive strength for 80 aluminum-lithium alloy specimens.

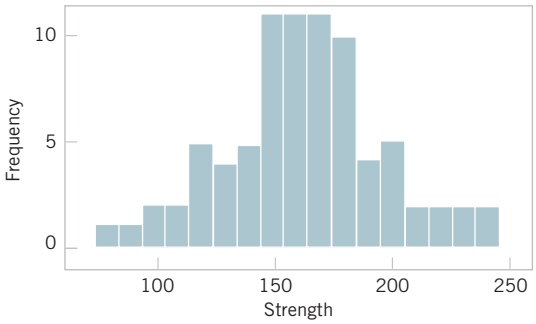


FIGURE 6.8

A histogram of the compressive strength data with 17 bins.

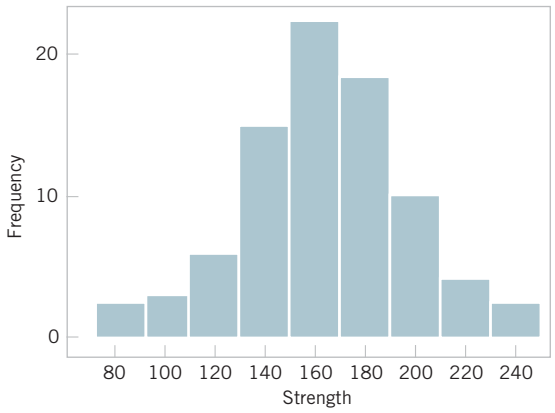


FIGURE 6.9

A histogram of the compressive strength data with nine bins.

Figure 6.8 is a histogram of the compressive strength data with 17 bins. We have noted that histograms may be relatively sensitive to the number of bins and their width. For small data sets, histograms may change dramatically in appearance if the number and/or width of the bins changes. Histograms are more stable and thus reliable for larger data sets, preferably of size 75 to 100 or more. Figure 6.9 is a histogram for the compressive strength data with nine bins. This is similar to the original histogram shown in Figure 6.7. Because the number of observations is moderately large ($n = 80$), the choice of the number of bins is not especially important, and both Figures 6.8 and 6.9 convey similar information.

Figure 6.10 is a variation of the histogram available in some software packages, the **cumulative frequency plot**. In this plot, the height of each bar is the total number of observations that are less than or equal to the upper limit of the bin. Cumulative distributions are also useful in data interpretation; for example, we can read directly from Figure 6.10 that approximately 70 observations are less than or equal to 200 psi.

When the sample size is large, the histogram can provide a reasonably reliable indicator of the general **shape** of the distribution or population of measurements from which the sample was drawn. See Figure 6.11 for three cases. The median is denoted as $\tilde{x}$. Generally, if the data are symmetric, as in Figure 6.11(b), the mean and median coincide. If, in addition, the data have only one mode (we say the data are *unimodal*), the mean, median, and mode all coincide.

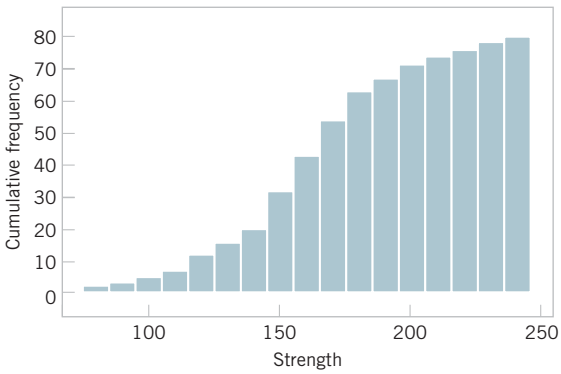


FIGURE 6.10

A cumulative distribution plot of the compressive strength data.

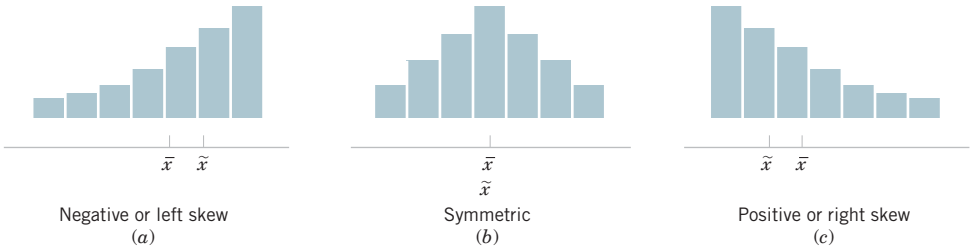


FIGURE 6.11

Histograms for symmetric and skewed distributions.

If the data are *skewed* (asymmetric, with a long tail to one side), as in Figure 6.11(a) and (c), the mean, median, and mode do not coincide. Usually, we find that mode < median < mean if the distribution is skewed to the right, whereas mode > median > mean if the distribution is skewed to the left.

Frequency distributions and histograms can also be used with qualitative or categorical data. Some applications will have a natural ordering of the categories (such as freshman, sophomore, junior, and senior), whereas in others, the order of the categories will be arbitrary (such as male and female). When using categorical data, the bins should have equal width.

EXAMPLE 6.6

Figure 6.12 presents the production of transport aircraft by the Boeing Company in 1985. Notice that the 737 was the most popular model, followed by the 757, 747, 767, and 707.

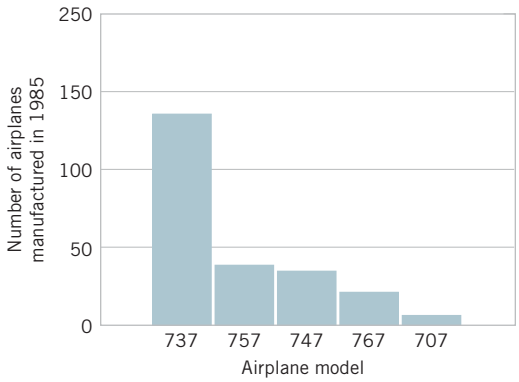


FIGURE 6.12

Airplane production in 1985.

A chart of occurrences by category (in which the categories are ordered by the number of occurrences) is sometimes referred to as a **Pareto chart**. An exercise asks you to construct such a chart.

In this section, we have concentrated on descriptive methods for the situation in which each observation in a data set is a single number or belongs to one category. In many cases, we work with data in which each observation consists of several measurements. For example, in a gasoline mileage study, each observation might consist of a measurement of miles per gallon, the size of the engine in the vehicle, engine horsepower, vehicle weight, and vehicle length. This is an example of **multivariate data**. In Section 6.6, we illustrate one simple graphical display for multivariate data. In later chapters, we discuss analyzing this type of data.

6.4 Box Plots

The stem-and-leaf display and the histogram provide general visual impressions about a data set, but numerical quantities such as $\bar{x}$ or s provide information about only one feature of the data. The **box plot** is a graphical display that simultaneously describes several important features of a data set, such as center, spread, departure from symmetry, and identification of unusual observations or outliers.

A box plot, sometimes called *box-and-whisker plots*, displays the three quartiles, the minimum, and the maximum of the data on a rectangular box, aligned either horizontally or vertically. The box encloses the interquartile range with the left (or lower) edge at the first quartile, q_1, and the right (or upper) edge at the third quartile, q_3. A line is drawn through the box at the second quartile (which is the 50th percentile or the median), $q_2 = \tilde{x}$. A line, or **whisker**, extends from each end of the box. The lower whisker is a line from the first quartile to the smallest data point within 1.5 interquartile ranges from the first quartile. The upper whisker is a line from the third quartile to the largest data point within 1.5 interquartile ranges from the third quartile. Data farther from the box than the whiskers are plotted as individual points. A point beyond a whisker, but less than three interquartile ranges from the box edge, is called an **outlier**. A point more than three interquartile ranges from the box edge is called an **extreme outlier**. See Figure 6.13. Occasionally, different symbols, such as open and filled circles, are used to identify the two types of outliers.

Figure 6.14 presents a typical computer-generated box plot for the alloy compressive strength data shown in Table 6.2. This box plot indicates that the distribution of compressive strengths is

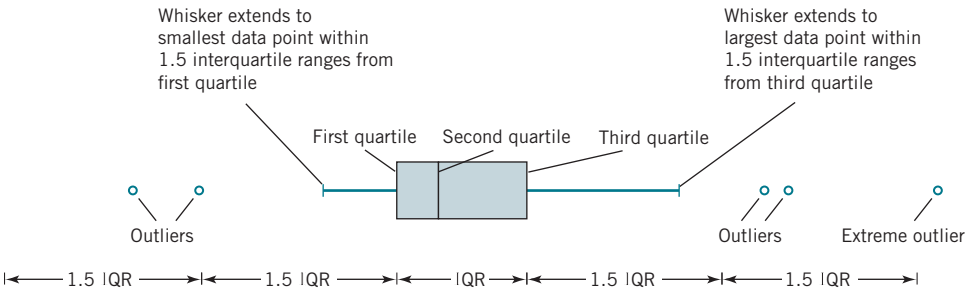


FIGURE 6.13

Description of a box plot.

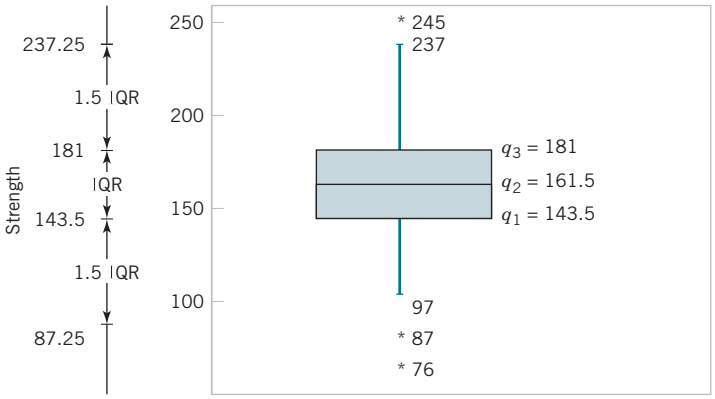


FIGURE 6.14

Box plot for compressive strength data in Table 6.2.

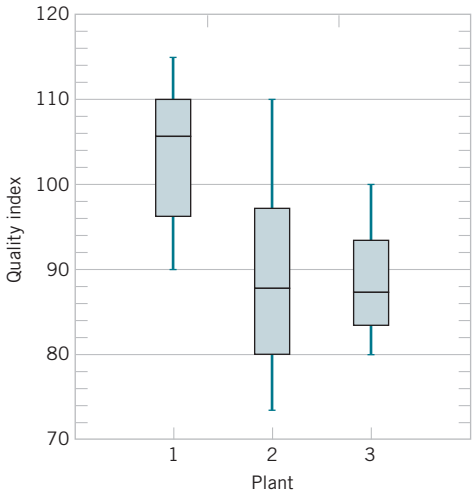


FIGURE 6.15

Comparative box plots of a quality index at three plants.

fairly symmetric around the central value because the left and right whiskers and the lengths of the left and right boxes around the median are about the same. There are also two mild outliers at lower strength and one at higher strength. The upper whisker extends to observation 237 because it is the highest observation below the limit for upper outliers. This limit is $q_3 + 1.5\text{IQR} = 181 + 1.5(181 - 143.5) = 237.25$. The lower whisker extends to observation 97 because it is the smallest observation above the limit for lower outliers. This limit is $q_1 - 1.5\text{IQR} = 143.5 - 1.5(181 - 143.5) = 87.25$. Box plots are very useful in graphical comparisons among data sets because they have high visual impact and are easy to understand. For example, Figure 6.15 shows the comparative box plots for a manufacturing quality index on semiconductor devices at three manufacturing plants. Inspection of this display reveals that there is too much variability at plant 2 and that plants 2 and 3 need to raise their quality index performance.

6.5 Time Sequence Plots

The graphical displays that we have considered thus far such as histograms, stem-and-leaf plots, and box plots are very useful visual methods for showing the variability in data. However, we noted in Chapter 1 that time is an important factor that contributes to variability in data, and those graphical methods do not take this into account. A **time series** or **time sequence** is a data set in which the observations are recorded in the order in which they occur. A **time series plot** is a graph in which the vertical axis denotes the observed value of the variable (say, x) and the horizontal axis denotes the time (which could be minutes, days, years, etc.). When measurements are plotted as a time series, we often see trends, cycles, or other broad features of the data that could not be seen otherwise.

For example, consider Figure 6.16(a), which presents a time series plot of the annual sales of a company for the last 10 years. The general impression from this display is that sales show an upward **trend**. There is some variability about this trend, with some years' sales increasing over those of the last year and some years' sales decreasing. Figure 6.16(b) shows the last 3 years of sales reported by quarter. This plot clearly shows that the annual sales in this business exhibit a **cyclic** variability by quarter, with the first- and second-quarter sales being generally higher than sales during the third and fourth quarters.

Sometimes it can be very helpful to combine a time series plot with some of the other graphical displays that we have considered previously. J. Stuart Hunter ("The Digidot Plot," *The American Statistician*, 1988, Vol. 42, p. 54) has suggested combining the stem-and-leaf plot with a time series plot to form a **digidot plot**.

Figure 6.17 is a digidot plot for the observations on compressive strength from Table 6.2, assuming that these observations are recorded in the order in which they occurred. This plot

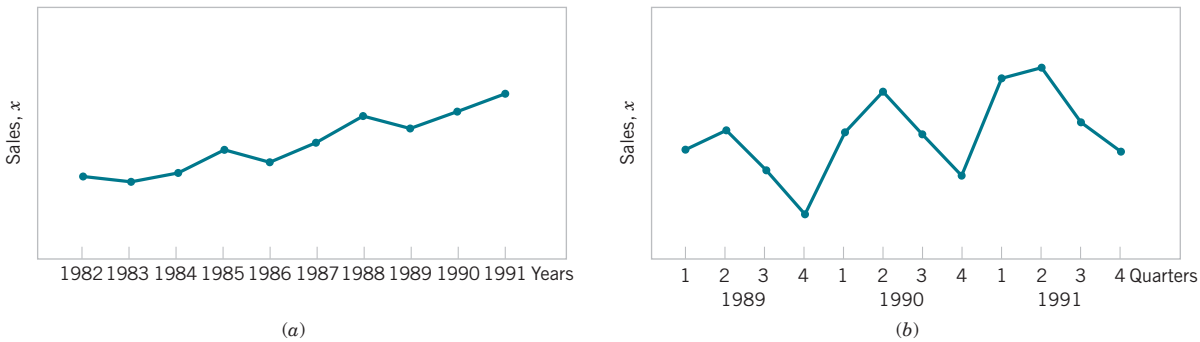


FIGURE 6.16

Company sales by year (*a*). By quarter (*b*).

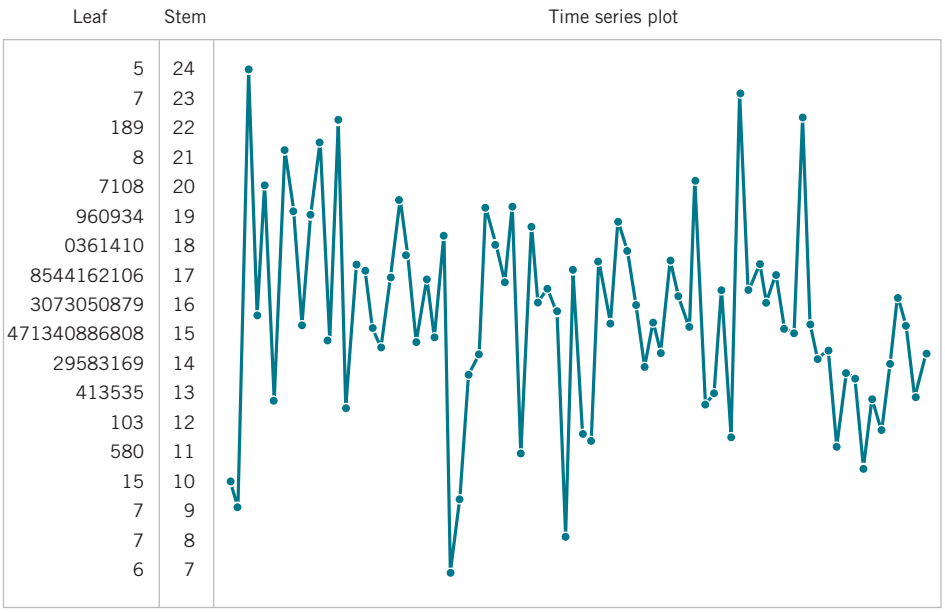


FIGURE 6.17

A digidot plot of the compressive strength data in Table 6.2.

effectively displays the overall variability in the compressive strength data and simultaneously shows the variability in these measurements over time. The general impression is that compressive strength varies around the mean value of 162.66, and no strong obvious pattern occurs in this variability over time.

The digidot plot in Figure 6.18 tells a different story. This plot summarizes 30 observations on concentration of the output product from a chemical process where the observations are recorded at one-hour time intervals. This plot indicates that during the first 20 hours of operation, this process produced concentrations generally above 85 grams per liter, but that following sample 20, something may have occurred in the process that resulted in lower concentrations. If this variability in output product concentration can be reduced, operation of this process can be improved. Notice that this apparent change in the process output is not seen in the stem-and-leaf portion of the digidot plot. The stem-and-leaf plot compresses the time dimension out of the data. This illustrates why it is always important to construct a time series plot for time-oriented data.

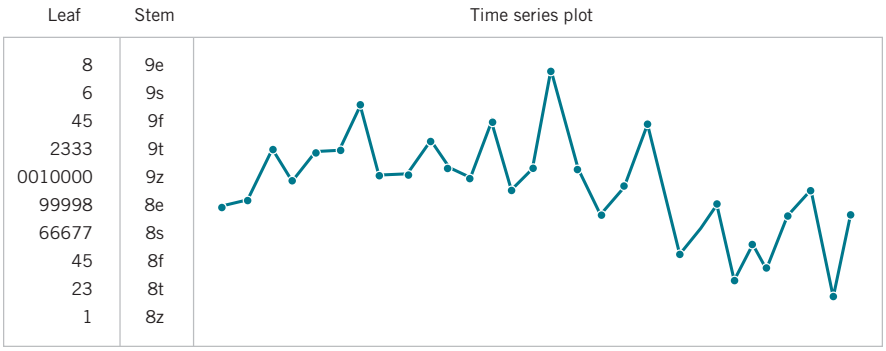


FIGURE 6.18

A digidot plot of chemical process concentration readings, observed hourly.

6.6 Scatter Diagrams

In many problems, engineers and scientists work with data that is **multivariate** in nature; that is, each observation consists of measurements of several variables. We saw an example of this in the wire bond pull strength data in Table 1.2. Each observation consisted of data on the pull strength of a particular wire bond, the wire length, and the die height. Such data are very commonly encountered. Table 6.5 contains a second example of multivariate data taken from an article on the quality of different young red wines in the *Journal of the Science of Food and Agriculture* (1974, Vol. 25(11), pp. 1369–1379) by T.C. Somers and M.E. Evans. The authors reported quality along with several other descriptive variables. We show only quality, pH, total SO_2 (in ppm), color density, and wine color for a sample of their wines.

Suppose that we wanted to graphically display the potential relationship between quality and one of the other variables, say color. The **scatter diagram** is a useful way to do this. Plot each pair of observations with one measurement in the pair on the vertical axis of the graph and the other measurement in the pair on the horizontal axis.

Figure 6.19 is the scatter diagram of quality versus the descriptive variable color. Notice that there is an apparent relationship between the two variables, with wines of more intense color generally having a higher quality rating.

TABLE 6.5 Quality Data for Young Red Wines

Quality	pH	Total SO_2	Color Density	Color
19.2	3.85	66	9.35	5.65
18.3	3.75	79	11.15	6.95
17.1	3.88	73	9.40	5.75
15.2	3.66	86	6.40	4.00
14.0	3.47	178	3.60	2.25
13.8	3.75	108	5.80	3.20
12.8	3.92	96	5.00	2.70
17.3	3.97	59	10.25	6.10
16.3	3.76	22	8.20	5.00
16.0	3.98	58	10.15	6.00
15.7	3.75	120	8.80	5.50
15.3	3.77	144	5.60	3.35
14.3	3.76	100	5.55	3.25
14.0	3.76	104	8.70	5.10
13.8	3.90	67	7.41	4.40
12.5	3.80	89	5.35	3.15
11.5	3.65	192	6.35	3.90
14.2	3.60	301	4.25	2.40
17.3	3.86	99	12.85	7.70
15.8	3.93	66	4.90	2.75

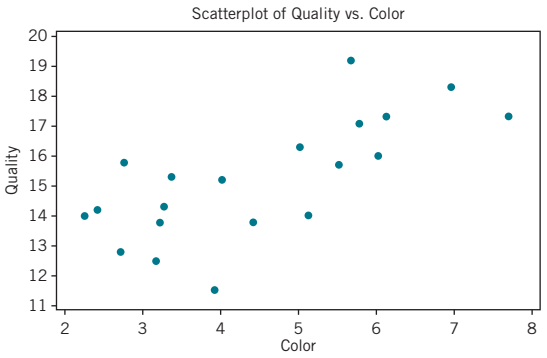


FIGURE 6.19

Scatter diagram of wine quality and color from Table 6.5.

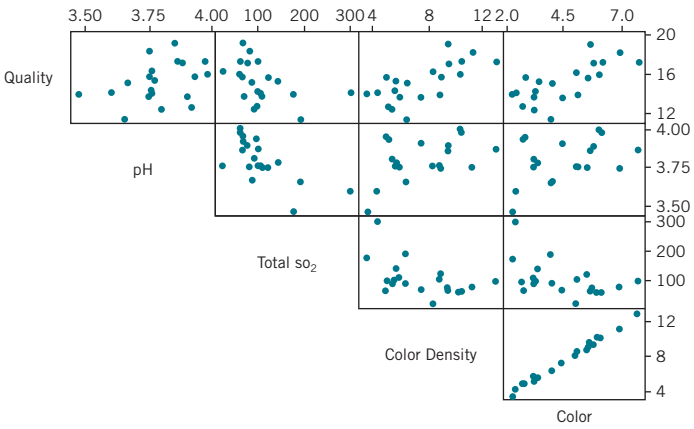


FIGURE 6.20

Matrix of scatter diagrams for the wine quality data in Table 6.5.

A scatter diagram is an excellent exploratory tool and can be very useful in identifying potential relationships between two variables. Data in Figure 6.19 indicate that a linear relationship between quality and color may exist. We saw an example of a three-dimensional scatter diagram in Chapter 1 where we plotted wire bond strength versus wire length and die height for the bond pull strength data.

When two or more variables exist, the **matrix of scatter diagrams** may be useful in looking at all of the pairwise relationships between the variables in the sample. Figure 6.20 is the matrix of scatter diagrams (upper half only shown) for the wine quality data in Table 6.5. The top row of the graph contains individual scatter diagrams of quality versus the other four descriptive variables, and other cells contain other pairwise plots of the four descriptive variables pH, SO_2, color density, and color. This display indicates a weak potential linear relationship between quality and pH and somewhat stronger potential relationships between quality and color density and quality and color (which was noted previously in Figure 6.19). A strong apparent linear relationship between color density and color exists (this should be expected).

The **sample correlation coefficient** r_{xy} is a quantitative measure of the strength of the linear relationship between two random variables x and y. The sample correlation coefficient is defined as

$$r_{xy} = \frac{\sum_{i=1}^{n} y_i(x_i - \bar{x})}{\left[\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{1/2}} \tag{6.6}$$

If the two variables are perfectly linearly related with a positive slope, then $r_{xy} = 1$, and if they are perfectly linearly related with a negative slope, then $r_{xy} = -1$. If no linear relationship between the two variables exists, then $r_{xy} = 0$. The simple correlation coefficient is also sometimes called the **Pearson correlation coefficient** after Karl Pearson, one of the giants of the fields of statistics in the late 19th and early 20th centuries.

The value of the sample correlation coefficient between quality and color, the two variables plotted in the scatter diagram of Figure 6.19, is 0.712. This is a moderately strong correlation, indicating a possible linear relationship between the two variables. Correlations below $|0.5|$ are generally considered weak and correlations above $|0.8|$ are generally considered strong.

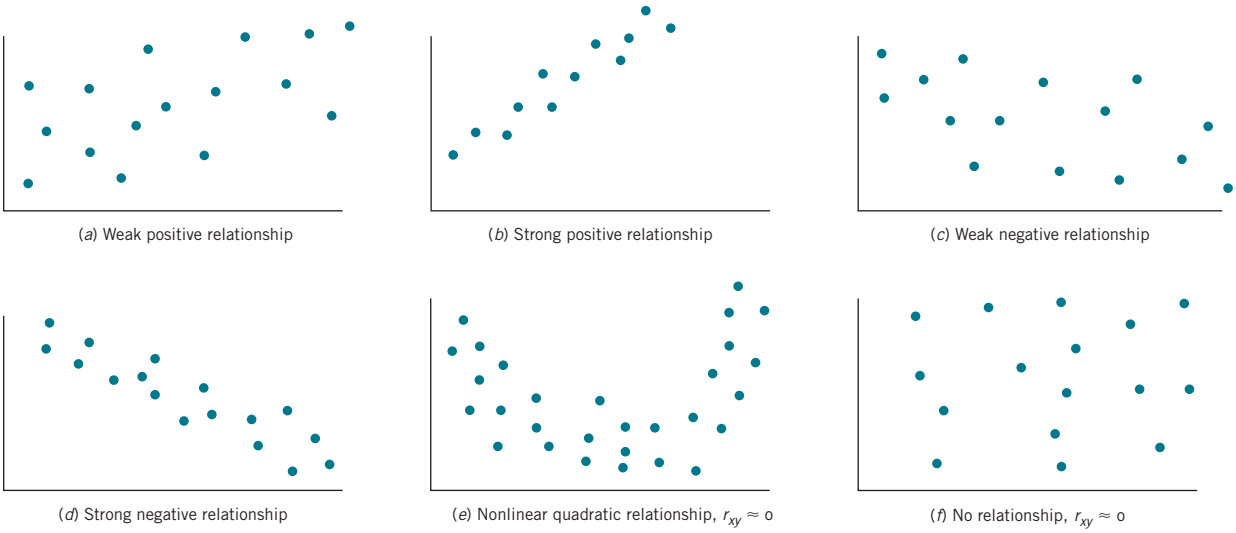

(a) Weak positive relationship
(b) Strong positive relationship
(c) Weak negative relationship
(d) Strong negative relationship
(e) Nonlinear quadratic relationship, $r_{xy} \approx 0$
(f) No relationship, $r_{xy} \approx 0$

FIGURE 6.21

Potential relationship between variables.

All pairwise sample correlations between the five variables in Table 6.5 are as follows:

Quality	pH	Total SO_2	Color	Density
pH	0.349			
Total SO_2	−0.445	−0.679		
Color density	0.702	0.482	−0.492	
Color	0.712	0.430	−0.480	0.996

Moderately strong correlations exist between quality and the two variables color and color density and between pH and total SO_2 (note that this correlation is negative). The correlation between color and color density is 0.996, indicating a nearly perfect linear relationship.

See Figure 6.21 for several examples of scatter diagrams exhibiting possible relationships between two variables. Parts (*e*) and (*f*) of the figure deserve special attention; in part (*e*), a probable quadratic relationship exists between y and x, but the sample correlation coefficient is close to zero because the correlation coefficient is a measure of **linear** association, but in part (*f*), the correlation is approximately zero because no association exists between the two variables.

6.7 Probability Plots

How do we know whether a particular probability distribution is a reasonable model for data? Sometimes this is an important question because many of the statistical techniques presented in subsequent chapters are based on an assumption that the population distribution is of a specific type. Thus, we can think of determining whether data come from a specific probability distribution as **verifying assumptions**. In other cases, the form of the distribution can give insight into the underlying physical mechanism generating the data. For example, in reliability engineering, verifying that time-to-failure data come from an exponential distribution identifies the **failure mechanism** in the sense that the failure rate is constant with respect to time.

Some of the visual displays we used earlier, such as the histogram, can provide insight about the form of the underlying distribution. However, histograms are usually not really reliable indicators of the distribution form unless the sample size is very large. A **probability plot** is a

graphical method for determining whether sample data conform to a hypothesized distribution based on a subjective visual examination of the data. The general procedure is very simple and can be performed quickly. It is also more reliable than the histogram for small- to moderate-size samples. Probability plotting typically uses special axes that have been scaled for the hypothesized distribution. Software is widely available for the normal, lognormal, Weibull, and various chi-square and gamma distributions. We focus primarily on normal probability plots because many statistical techniques are appropriate only when the population is (at least approximately) normal.

To construct a probability plot, the observations in the sample are first ranked from smallest to largest. That is, the sample $x_1, x_2, \ldots, x_n$ is arranged as $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, where $x_{(1)}$ is the smallest observation, $x_{(2)}$ is the second-smallest observation, and so forth with $x_{(n)}$ the largest. The ordered observations $x_{(j)}$ are then plotted against their observed cumulative frequency $(j - 0.5)/n$ on the appropriate probability paper. If the hypothesized distribution adequately describes the data, the plotted points will fall approximately along a straight line; if the plotted points deviate significantly from a straight line, the hypothesized model is not appropriate. Usually, the determination of whether or not the data plot is a straight line is subjective. The procedure is illustrated in the following example.

EXAMPLE 6.7 | Battery Life

Ten observations on the effective service life in minutes of batteries used in a portable personal computer are as follows: 176, 191, 214, 220, 205, 192, 201, 190, 183, 185. We hypothesize that battery life is adequately modeled by a normal distribution. To use probability plotting to investigate this hypothesis, first arrange the observations in ascending order and calculate their cumulative frequencies $(j - 0.5)/10$ as shown in Table 6.6.

TABLE 6.6 **Calculation for Constructing a Normal Probability Plot**

j	$x_{(j)}$	$(j - 0.5)/10$	z_j
1	176	0.05	−1.64
2	183	0.15	−1.04
3	185	0.25	−0.67
4	190	0.35	−0.39
5	191	0.45	−0.13
6	192	0.55	0.13
7	201	0.65	0.39
8	205	0.75	0.67
9	214	0.85	1.04
10	220	0.95	1.64

The pairs of values $x_{(j)}$ and $(j - 0.5)/10$ are now plotted on normal probability axes. This plot is shown in Figure 6.22.

Most normal probability plots have $100(j - 0.5)/n$ on the left vertical scale and (sometimes) $100[1 - (j - 0.5)/n]$ on the right vertical scale, with the variable value plotted on the horizontal scale. A straight line, chosen subjectively, has been drawn through the plotted points. In drawing the straight line, you should be influenced more by the points near the middle of the plot than by the extreme points. A good rule of thumb is to draw the line approximately between the 25th and 75th percentile points. This is how the line in Figure 6.22 was determined. In assessing the "closeness" of the points to the straight line, imagine a "fat pencil" lying along the line. If all the points are covered by this imaginary pencil, a normal distribution adequately describes the data. Because the points in Figure 6.19 would pass the "fat pencil" test, we conclude that the normal distribution is an appropriate model.

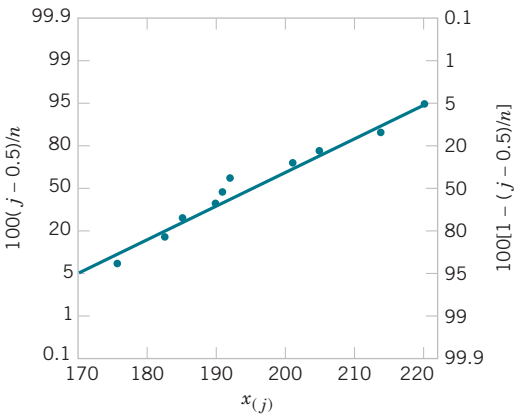


FIGURE 6.22
Normal probability plot for battery life.

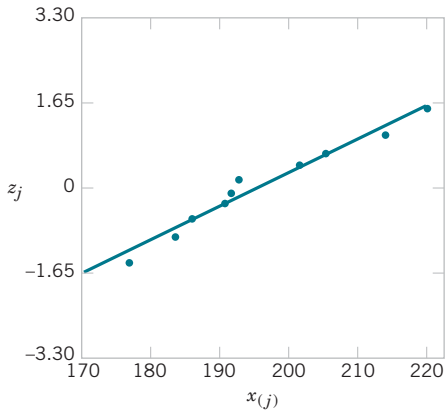


FIGURE 6.23
Normal probability plot obtained from standardized normal scores.

A **normal probability plot** can also be constructed on ordinary axes by plotting the standardized normal scores z_j against $x_{(j)}$, where the standardized normal scores satisfy

$$\frac{j - 0.5}{n} = P(Z \leq z_j) = \Phi(z_j)$$

For example, if $(j - 0.5)/n = 0.05$, $\Phi(z_j) = 0.05$ implies that $z_j = -1.64$. To illustrate, consider the data from Example 6.4. In the last column of Table 6.6 we show the standardized normal scores. Figure 6.23 is the plot of z_j versus $x_{(j)}$. This normal probability plot is equivalent to the one in Figure 6.22.

We have constructed our probability plots with the probability scale (or the z-scale) on the vertical axis. Some computer packages "flip" the axis and put the probability scale on the horizontal axis.

The normal probability plot can be useful in identifying distributions that are symmetric but that have tails that are "heavier" or "lighter" than the normal. They can also be useful in identifying skewed distributions. When a sample is selected from a light-tailed distribution (such as the uniform distribution), the smallest and largest observations will not be as extreme as would be expected in a sample from a normal distribution. Thus, if we consider the straight line drawn through the observations at the center of the normal probability plot, observations on the left side will tend to fall below the line, and observations on the right side will tend to fall above the line. This will produce an S-shaped normal probability plot such as shown in Figure 6.24(a). A heavy-tailed distribution will result in data that also produce an S-shaped normal probability plot, but now the observations on the left will be above the straight line and the observations on the right will lie below the line. See Figure 6.24(b). A positively skewed distribution will tend to produce a pattern such as shown in Figure 6.24(c), where points on both ends of the plot tend to fall below the line, giving a curved shape to the plot. This occurs because both the smallest and the largest observations from this type of distribution are larger than expected in a sample from a normal distribution.

Even when the underlying population is exactly normal, the sample data will not plot exactly on a straight line. Some judgment and experience are required to evaluate the plot. Generally, if the sample size is $n < 30$, there can be significant deviation from linearity in normal plots, so in these cases only a very severe departure from linearity should be interpreted as a strong indication of nonnormality. As n increases, the linear pattern will tend to become stronger, and the normal probability plot will be easier to interpret and more reliable as an indicator of the form of the distribution.

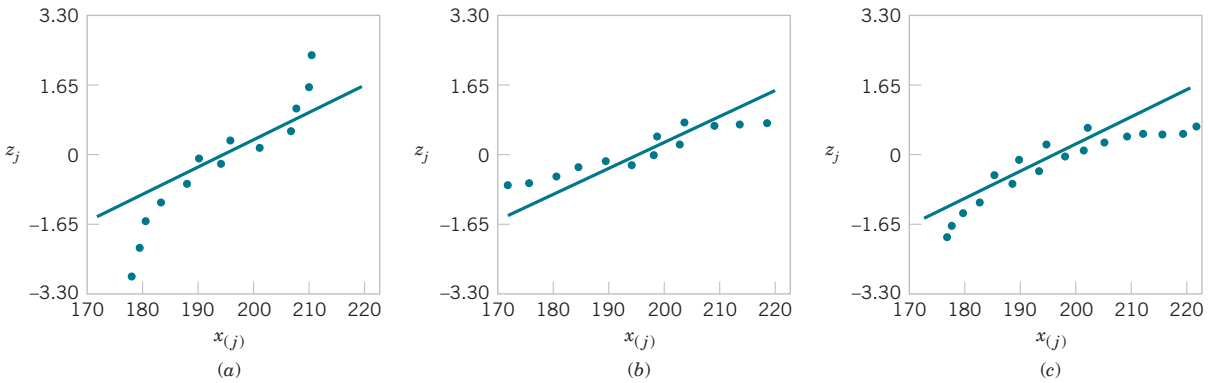


FIGURE 6.24

Normal probability plots indicating a nonnormal distribution. (a) Light-tailed distribution. (b) Heavy-tailed distribution. (c) A distribution with positive (or right) skew.

Important Terms and Concepts

Box plot
Degrees of freedom
Digidot plot
Frequency distribution and histogram
Histogram
Interquartile range
Matrix of scatter plots
Multivariate data
Normal probability plot
Outlier
Pareto chart
Percentile
Population mean
Population standard deviation
Population variance
Probability plot
Quartiles and percentiles
Relative frequency distribution
Sample correlation coefficient
Sample mean
Sample median
Sample mode
Sample range
Sample standard deviation
Sample variance
Scatter diagram
Stem-and-leaf diagram
Time series

CHAPTER 7

Point Estimation of Parameters and Sampling Distributions

© EasyBuy4u / iStockphoto

LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Explain the general concepts of estimating the parameters of a population or a probability distribution
2. Explain the important role of the normal distribution as a sampling distribution and the central limit theorem
3. Explain important properties of point estimators, including bias, variance, and mean square error
4. Construct point estimators using the method of moments and the method of maximum likelihood
5. Compute and explain the precision with which a parameter is estimated
6. Construct a point estimator from the Bayesian approach

CHAPTER OUTLINE

7.1 Point Estimation

7.2 Sampling Distributions and the Central Limit Theorem

7.3 General Concepts of Point Estimation

- **7.3.1** Unbiased Estimators
- **7.3.2** Variance of a Point Estimator
- **7.3.3** Standard Error: Reporting a Point Estimate
- **7.3.4** Bootstrap Standard Error
- **7.3.5** Mean Squared Error of an Estimator

7.4 Methods of Point Estimation

- **7.4.1** Method of Moments
- **7.4.2** Method of Maximum Likelihood
- **7.4.3** Bayesian Estimation of Parameters

Introduction

Statistical methods are used to make decisions and draw conclusions about populations. This aspect of statistics is generally called **statistical inference**. These techniques utilize the information in a sample for drawing conclusions. This chapter begins our study of the statistical methods used in decision making.

Statistical inference may be divided into two major areas: **parameter estimation** and **hypothesis testing**. As an example of a parameter estimation problem, suppose that an engineer is analyzing the tensile strength of a component used in an air frame. This is an important part of assessing the overall structural integrity of the airplane. Variability is naturally present in the individual components because of differences in the batches of raw material used to make the components, manufacturing processes, and measurement procedures (for example), so the engineer wants to estimate the mean strength of the population of components. In practice, the engineer will use sample data to compute a number that is in some sense a reasonable value (a good guess) of the true population mean. This number is called a **point estimate**. We will see that procedures are available for developing point estimates of parameters that have good statistical properties. We will also be able to establish the precision of the point estimate.

Now let's consider a different type of question. Suppose that two different reaction temperatures t_1 and t_2 can be used in a chemical process. The engineer conjectures that t_1 will result in higher yields than t_2. If the engineers can demonstrate that t_1 results in higher yields, then a process change can probably be justified. Statistical hypothesis testing is the framework for solving problems of this type. In this example, the engineer would be interested in formulating hypotheses that allow him or her to demonstrate that the mean yield using t_1 is higher than the mean yield using t_2. Notice that there is no emphasis on estimating yields; instead, the focus is on drawing conclusions about a hypothesis that is relevant to the engineering decision.

This chapter and Chapter 8 discuss parameter estimation. Chapters 9 and 10 focus on hypothesis testing.

7.1 Point Estimation

Statistical inference always focuses on drawing conclusions about one or more parameters of a population. An important part of this process is obtaining estimates of the parameters. Suppose that we want to obtain a point estimate (a reasonable value) of a population parameter. We know that before the data are collected, the observations are considered to be random variables, say, X_1, $X_2, \ldots, X_n$. Therefore, any function of the observations, or any **statistic**, is also a random variable. For example, the sample mean $\overline{X}$ and the sample variance S^2 are statistics and random variables.

A simple way to visualize this is as follows. Suppose we take a sample of $n = 10$ observations from a population and compute the sample average, getting the result $\overline{x} = 10.2$. Now we repeat this process, taking a second sample of $n = 10$ observations from the same population and the resulting sample average is 10.4. The sample average depends on the observations in the sample, which differ from sample to sample because they are random variables. Consequently, the sample average (or any other function of the sample data) is a random variable.

Because a statistic is a random variable, it has a probability distribution. We call the probability distribution of a statistic a **sampling distribution**. The notion of a sampling distribution is very important and is discussed and illustrated later in the chapter.

When discussing inference problems, it is convenient to have a general symbol to represent the parameter of interest. We use the Greek symbol θ (theta) to represent the parameter. The symbol θ can represent the mean μ, the variance σ^2, or any parameter of interest to us. The objective of point estimation is to select a single number based on sample data that is the most plausible value for θ. The numerical value of a sample statistic is used as the point estimate.

In general, if X is a random variable with probability distribution $f(x)$, characterized by the unknown parameter θ, and if $X_1, X_2, \ldots, X_n$ is a random sample of size n from X, the statistic

$\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$ is called a **point estimator** of θ. Note that $\hat{\Theta}$ is a random variable because it is a function of random variables. After the sample has been selected, $\hat{\Theta}$ takes on a particular numerical value $\hat{\theta}$ called the **point estimate** of θ.

Point Estimator

A **point estimate** of some population parameter θ is a single numerical value $\hat{\theta}$ of a statistic $\hat{\Theta}$. The statistic $\hat{\Theta}$ is called the **point estimator**.

As an example, suppose that the random variable X is normally distributed with an unknown mean μ. The sample mean is a point estimator of the unknown population mean μ. That is, $\hat{\mu} = \overline{X}$. After the sample has been selected, the numerical value $\overline{x}$ is the point estimate of μ. Thus, if $x_1 = 25$, $x_2 = 30$, $x_3 = 29$, and $x_4 = 31$, the point estimate of μ is

$$\overline{x} = \frac{25 + 30 + 29 + 31}{4} = 28.75$$

Similarly, if the population variance σ^2 is also unknown, a point estimator for σ^2 is the sample variance S^2, and the numerical value $s^2 = 6.9$ calculated from the sample data is called the *point estimate of* σ^2.

Estimation problems occur frequently in engineering. We often need to estimate

- The mean μ of a single population
- The variance σ^2 (or standard deviation σ) of a single population
- The proportion p of items in a population that belong to a class of interest
- The difference in means of two populations, $\mu_1 - \mu_2$
- The difference in two population proportions, $p_1 - p_2$

Reasonable point estimates of these parameters are as follows:

- For μ, the estimate is $\hat{\mu} = \overline{x}$, the sample mean.
- For σ^2, the estimate is $\hat{\sigma}^2 = s^2$, the sample variance.
- For p, the estimate is $\hat{p} = x/n$, the sample proportion, where x is the number of items in a random sample of size n that belong to the class of interest.
- For $\mu_1 - \mu_2$, the estimate is $\hat{\mu}_1 - \hat{\mu}_2 = \overline{x}_1 - \overline{x}_2$, the difference between the sample means of two independent random samples.
- For $p_1 - p_2$, the estimate is $\hat{p}_1 - \hat{p}_2$, the difference between two sample proportions computed from two independent random samples.

We may have several different choices for the point estimator of a parameter. For example, if we wish to estimate the mean of a population, we might consider the sample mean, the sample median, or perhaps the average of the smallest and largest observations in the sample as point estimators. To decide which point estimator of a particular parameter is the best one to use, we need to examine their statistical properties and develop some criteria for comparing estimators.

7.2 Sampling Distributions and the Central Limit Theorem

Statistical inference is concerned with making **decisions** about a population based on the information contained in a random sample from that population. For instance, we may be interested in the mean fill volume of a container of soft drink. The mean fill volume in the population is required to

be 300 milliliters. An engineer takes a random sample of 25 containers and computes the sample average fill volume to be $\bar{x} = 298.8$ milliliters. The engineer will probably decide that the population mean is $\mu = 300$ milliliters even though the sample mean was 298.8 milliliters because he or she knows that the sample mean is a reasonable estimate of μ and that a sample mean of 298.8 milliliters is very likely to occur even if the true population mean is $\mu = 300$ milliliters. In fact, if the true mean is 300 milliliters, tests of 25 containers made repeatedly, perhaps every 5 minutes, would produce values of $\bar{x}$ that vary both above and below $\mu = 300$ milliliters.

The link between the probability models in the earlier chapters and the data is made as follows. Each numerical value in the data is the observed value of a random variable. Furthermore, the random variables are usually assumed to be independent and identically distributed. These random variables are known as a *random sample*.

Random Sample

The random variables $X_1, X_2, \ldots, X_n$ are a **random sample** of size n if (a) the X_i's are independent random variables and (b) every X_i has the same probability distribution.

The observed data are also referred to as a *random sample*, but the use of the same phrase should not cause any confusion.

The assumption of a random sample is extremely important. If the sample is not random and is based on judgment or is flawed in some other way, statistical methods will not work properly and will lead to incorrect decisions.

The primary purpose in taking a random sample is to obtain information about the unknown population parameters. Suppose, for example, that we wish to reach a conclusion about the proportion of people in the United States who prefer a particular brand of soft drink. Let p represent the unknown value of this proportion. It is impractical to question every individual in the population to determine the true value of p. To make an inference regarding the true proportion p, a more reasonable procedure would be to select a random sample (of an appropriate size) and use the observed proportion $\hat{p}$ of people in this sample favoring the brand of soft drink.

The sample proportion, $\hat{p}$, is computed by dividing the number of individuals in the sample who prefer the brand of soft drink by the total sample size n. Thus, $\hat{p}$ is a function of the observed values in the random sample. Because many random samples are possible from a population, the value of $\hat{p}$ will vary from sample to sample. That is, $\hat{p}$ is a random variable. Such a random variable is called a **statistic**.

Statistic

A **statistic** is any function of the observations in a random sample.

We have encountered statistics before. For example, if $X_1, X_2, \ldots, X_n$ is a random sample of size n, the **sample mean** $\overline{X}$, the **sample variance** S^2, and the **sample standard deviation** S are statistics. Because a statistic is a random variable, it has a probability distribution.

Sampling Distribution

The probability distribution of a statistic is called a **sampling distribution**.

For example, the probability distribution of $\overline{X}$ is called the **sampling distribution of the mean**. The sampling distribution of a statistic depends on the distribution of the population,

the size of the sample, and the method of sample selection. We now present perhaps the most important sampling distribution. Other sampling distributions and their applications are illustrated extensively in the following two chapters.

Consider determining the sampling distribution of the sample mean $\overline{X}$. Suppose that a random sample of size n is taken from a normal population with mean μ and variance σ^2. Now each observation in this sample, say, $X_1, X_2, \ldots, X_n$, is a normally and independently distributed random variable with mean μ and variance σ^2. Then because linear functions of independent, normally distributed random variables are also normally distributed (Chapter 5), we conclude that the sample mean

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

has a normal distribution with mean

$$\mu_{\overline{X}} = \frac{\mu + \mu + \cdots + \mu}{n} = \mu$$

and variance

$$\sigma^2_{\overline{X}} = \frac{\sigma^2 + \sigma^2 + \cdots + \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

If we are sampling from a population that has an unknown probability distribution, the sampling distribution of the sample mean will still be approximately normal with mean μ and variance σ^2/n if the sample size n is large. This is one of the most useful theorems in statistics, called the **central limit theorem**. The statement is as follows:

Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population (either finite or infinite) with mean μ and finite variance σ^2 and if $\overline{X}$ is the sample mean, the limiting form of the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{7.1}$$

as $n \to \infty$, is the standard normal distribution.

It is easy to demonstrate the central limit theorem with a **computer simulation experiment**. Consider the lognormal distribution in Figure 7.1. This distribution has parameters $\theta = 2$ (called the **location** parameter) and $\omega = 0.75$ (called the **scale** parameter), resulting in mean $\mu = 9.79$ and standard deviation $\sigma = 8.51$. Notice that this lognormal distribution does not look very much like the normal distribution; it is defined only for positive values of the random variable X and is skewed considerably to the right. We used computer software to draw 20 samples at random from this distribution, each of size $n = 10$. The data from this sampling experiment are shown in Table 7.1. The last row in this table is the average of each sample $\overline{x}$.

The first thing that we notice in looking at the values of $\overline{x}$ is that they are not all the same. This is a clear demonstration of the point made previously that any statistic is a random variable. If we had calculated any sample statistic (s, the sample median, the upper or lower quartile, or a percentile), they would also have varied from sample to sample because they are random variables. Try it and see for yourself.

According to the central limit theorem, the distribution of the sample average $\overline{X}$ is normal. Figure 7.2 is a normal probability plot of the 20 sample averages $\overline{x}$ from Table 7.1. The observations scatter generally along a straight line, providing evidence that the distribution of the sample mean is normal even though the distribution of the population is very nonnormal. This type of sampling experiment can be used to investigate the sampling distribution of any statistic.

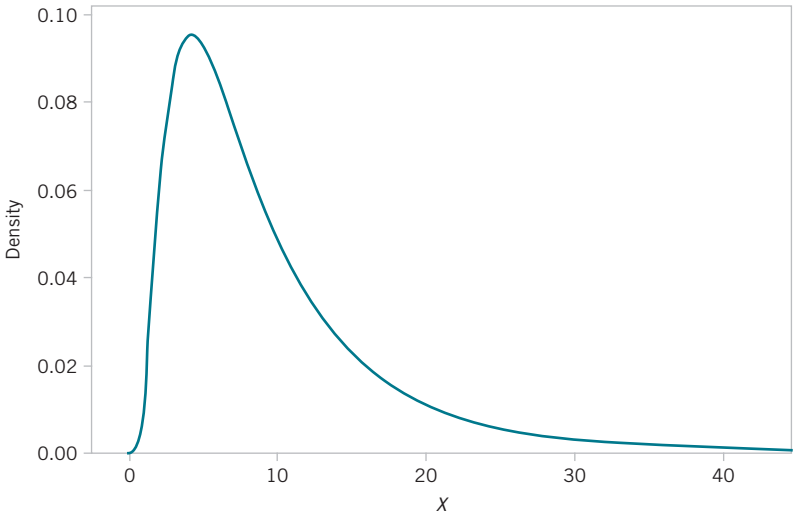


FIGURE 7.1

A lognormal distribution with θ = 2 and ω = 0.75.

TABLE 7.1 **Twenty samples of size $n = 10$ from the lognormal distribution in Figure 7.1.**

	Sample									
Obs	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	3.9950	8.2220	4.1893	15.0907	12.8233	15.2285	5.6319	7.5504	2.1503	3.1390
2	7.8452	13.8194	2.6186	4.5107	3.1392	16.3821	3.3469	1.4393	46.3631	1.8314
3	1.8858	4.0513	8.7829	7.1955	7.1819	12.0456	8.1139	6.0995	2.4787	3.7612
4	16.3041	7.5223	2.5766	18.9189	4.2923	13.4837	13.6444	8.0837	19.7610	15.7647
5	9.7061	6.7623	4.4940	11.1338	3.1460	13.7345	9.3532	2.1988	3.8142	3.6519
6	7.6146	5.3355	10.8979	3.6718	21.1501	1.6469	4.9919	13.6334	2.8456	14.5579
7	6.2978	6.7051	6.0570	8.5411	3.9089	11.0555	6.2107	7.9361	11.4422	9.7823
8	19.3613	15.6610	10.9201	5.9469	8.5416	19.7158	11.3562	3.9083	12.8958	2.2788
9	7.2275	3.7706	38.3312	6.0463	10.1081	2.2129	11.2097	3.7184	28.2844	26.0186
10	16.2093	3.4991	6.6584	4.2594	6.1328	9.2619	4.1761	5.2093	10.0632	17.9411
$\bar{x}$	9.6447	7.5348	9.5526	8.5315	8.0424	11.4767	7.8035	5.9777	14.0098	9.8727
Obs	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	7.5528	8.4998	2.5299	2.3115	6.1115	3.9102	2.3593	9.6420	5.0707	6.8075
2	4.9644	3.9780	11.0097	18.8265	3.1343	11.0269	7.3140	37.4338	5.5860	8.7372
3	16.7181	6.2696	21.9326	7.9053	2.3187	12.0887	5.1996	3.6109	3.6879	19.2486
4	8.2167	8.1599	15.5126	7.4145	6.7088	8.3312	11.9890	11.0013	5.6657	5.3550
5	9.0399	15.9189	7.9941	22.9887	8.0867	2.7181	5.7980	4.4095	12.1895	16.9185
6	4.0417	2.8099	7.1098	1.4794	14.5747	8.6157	7.8752	7.5667	32.7319	8.2588
7	4.9550	40.1865	5.1538	8.1568	4.8331	14.4199	4.3802	33.0634	11.9011	4.8917
8	7.5029	10.1408	2.6880	1.5977	7.2705	5.8623	2.0234	6.4656	12.8903	3.3929
9	8.4102	6.4106	7.6495	7.2551	3.9539	16.4997	1.8237	8.1360	7.4377	15.2643
10	7.2316	11.5961	4.4851	23.0760	10.3469	9.9330	8.6515	1.6852	3.6678	2.9765
$\bar{x}$	7.8633	11.3970	8.6065	10.1011	6.7339	9.3406	5.7415	12.3014	10.0828	9.1851

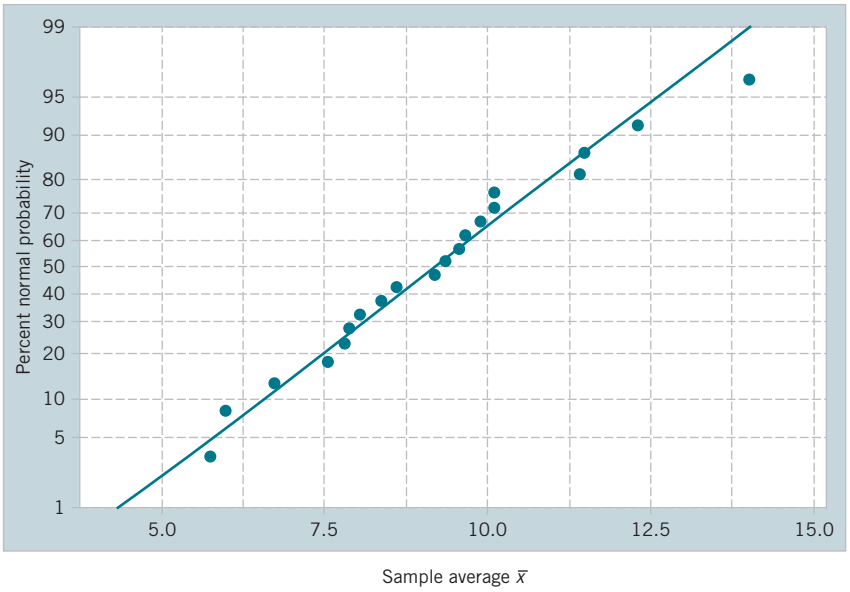


FIGURE 7.2

Normal probability plot of the sample averages from Table 7.1.

The normal approximation for $\overline{X}$ depends on the sample size n. Figure 7.3(a) is the distribution obtained for throws of a single, six-sided true die. The probabilities are equal (1/6) for all the values obtained: 1, 2, 3, 4, 5, or 6. Figure 7.3(b) is the distribution of the average score obtained when tossing two dice, and Figure 7.3(c), 7.3(d), and 7.3(e) show the distributions of average scores obtained when tossing 3, 5, and 10 dice, respectively. Notice that, although the population (one die) is relatively far from normal, the distribution of averages is approximated reasonably well by the normal distribution for sample sizes as small as five. (The dice throw distributions are discrete, but the normal is continuous.)

The central limit theorem is the underlying reason why many of the random variables encountered in engineering and science are normally distributed. The observed variable results from a series of underlying disturbances that act together to create a central limit effect.

When is the sample size large enough so that the central limit theorem can be assumed to apply? The answer depends on how close the underlying distribution is to the normal. If the underlying distribution is symmetric and unimodal (not too far from normal), the central limit theorem will apply for small values of n, say 4 or 5. If the sampled population is very nonnormal, larger samples will be required. As a general guideline, if $n > 30$, the central limit theorem will almost always apply. There are exceptions to this guideline, but they are relatively rare. In most

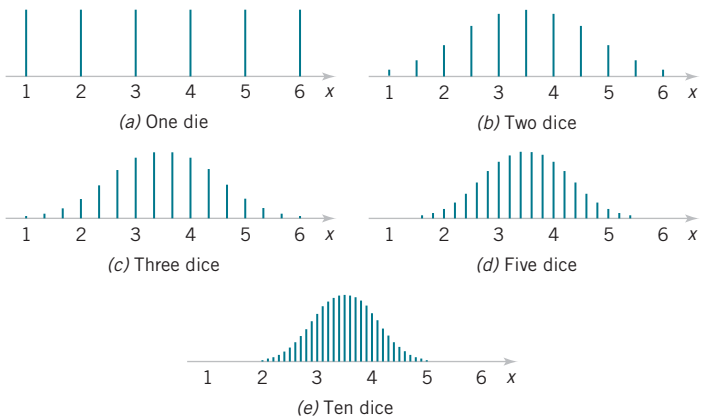


FIGURE 7.3

Distributions of average scores from throwing dice.

Source: Adapted with permission from Box, Hunter, and Hunter (1978).

cases encountered in practice, this guideline is very conservative, and the central limit theorem will apply for sample sizes much smaller than 30. For example, consider the dice example in Figure 7.3.

EXAMPLE 7.1 | Resistors

An electronics company manufactures resistors that have a mean resistance of 100 ohms and a standard deviation of 10 ohms. The distribution of resistance is normal. Find the probability that a random sample of $n = 25$ resistors will have an average resistance of fewer than 95 ohms.

Note that the sampling distribution of $\overline{X}$ is normal with mean $\mu_{\overline{X}} = 100$ ohms and a standard deviation of

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

Therefore, the desired probability corresponds to the shaded area in Figure 7.4. Standardizing point $\overline{X} = 95$ in Figure 7.4, we find that

$$z = \frac{95 - 100}{2} = -2.5$$

and therefore,

$$P(\overline{X} < 95) = P(Z < -2.5) = 0.0062$$

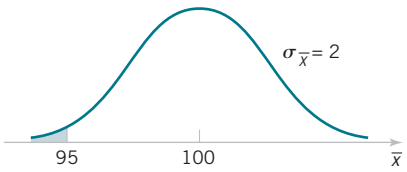


FIGURE 7.4

Probability for Example 7.1.

Practical Conclusion: This example shows that if the distribution of resistance is normal with mean 100 ohms and standard deviation of 10 ohms, finding a random sample of resistors with a sample mean less than 95 ohms is a **rare event**. If this actually happens, it casts doubt as to whether the true mean is really 100 ohms or if the true standard deviation is really 10 ohms.

The following example makes use of the central limit theorem.

EXAMPLE 7.2 | Central Limit Theorem

Suppose that a random variable X has a continuous uniform distribution

$$f(x) = \begin{cases} 1/2, & 4 \le x \le 6 \\ 0, & \text{otherwise} \end{cases}$$

Find the distribution of the sample mean of a random sample of size $n = 40$.

The mean and variance of X are $\mu = 5$ and $\sigma^2 = (6-4)^2/12 = 1/3$. The central limit theorem indicates that the distribution of $\overline{X}$ is approximately normal with mean $\mu_{\overline{X}} = 5$ and variance $\sigma^2_{\overline{X}} = \sigma^2/n = 1/[3(40)] = 1/120$. See the distributions of X and $\overline{X}$ in Figure 7.5.

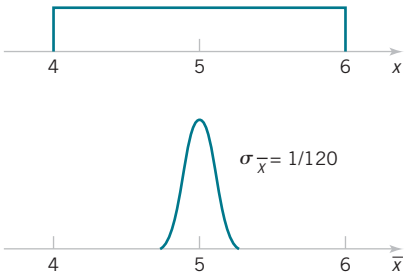


FIGURE 7.5

The distribution of X and $\overline{X}$ for Example 7.2.

Now consider the case in which we have two independent populations. Let the first population have mean μ_1 and variance σ_1^2 and the second population have mean μ_2 and variance σ_2^2. Suppose that both populations are normally distributed. Then, using the fact that linear combinations of independent normal random variables follow a normal distribution (see Chapter 5), we can say that the sampling distribution of $\overline{X}_1 - \overline{X}_2$ is normal with mean

$$\mu_{\overline{X}_1 - \overline{X}_2} = \mu_{\overline{X}_1} - \mu_{\overline{X}_2} = \mu_1 - \mu_2 \tag{7.2}$$

and variance

$$\sigma^2_{\overline{X}_1 - \overline{X}_2} = \sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \tag{7.3}$$

If the two populations are not normally distributed and if both sample sizes n_1 and n_2 are more than 30, we may use the central limit theorem and assume that $\overline{X}_1$ and $\overline{X}_2$ follow approximately independent normal distributions. Therefore, the sampling distribution of $\overline{X}_1 - \overline{X}_2$ is approximately normal with mean and variance given by Equations 7.2 and 7.3, respectively. If either n_1 or n_2 is fewer than 30, the sampling distribution of $\overline{X}_1 - \overline{X}_2$ will still be approximately normal with mean and variance given by Equations 7.2 and 7.3 provided that the population from which the small sample is taken is not dramatically different from the normal. We may summarize this with the following definition.

Approximate Sampling Distribution of a Difference in Sample Means

If we have two independent populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 and if $\overline{X}_1$ and $\overline{X}_2$ are the sample means of two independent random samples of sizes n_1 and n_2 from these populations, then the sampling distribution of

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \tag{7.4}$$

is approximately standard normal if the conditions of the central limit theorem apply. If the two populations are normal, the sampling distribution of Z is exactly standard normal.

7.3 General Concepts of Point Estimation

7.3.1 Unbiased Estimators

An estimator should be "close" in some sense to the true value of the unknown parameter. Formally, we say that $\hat{\Theta}$ is an unbiased estimator of θ if the expected value of $\hat{\Theta}$ is equal to θ. This is equivalent to saying that the mean of the probability distribution of $\hat{\Theta}$ (or the mean of the sampling distribution of $\hat{\Theta}$) is equal to θ.

Bias of an Estimator

The point estimator $\hat{\Theta}$ is an **unbiased estimator** for the parameter θ if

$$E(\hat{\Theta}) = \theta \tag{7.5}$$

If the estimator is not unbiased, then the difference

$$E(\hat{\Theta}) - \theta \tag{7.6}$$

is called the **bias** of the estimator $\hat{\Theta}$.

When an estimator is unbiased, the bias is zero; that is, $E(\hat{\Theta}) - \theta = 0$.

EXAMPLE 7.3 | Sample Mean and Variance Are Unbiased

Suppose that X is a random variable with mean μ and variance σ^2. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the population represented by X. Show that the sample mean $\overline{X}$ and sample variance S^2 are unbiased estimators of μ and σ^2, respectively.

First consider the sample mean. In Section 5.5 in Chapter 5, we showed that $E(\overline{X}) = \mu$. Therefore, the sample mean $\overline{X}$ is an unbiased estimator of the population mean μ.

Now consider the sample variance. We have

$$E(S^2) = E\left[\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}\right] = \frac{1}{n-1}E\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$= \frac{1}{n-1}E\sum_{i=1}^{n}(X_i^2 + \overline{X}^2 - 2\overline{X}X_i)$$

$$= \frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2 - n\overline{X}^2\right)$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}E(X_i^2) - nE(\overline{X}^2)\right]$$

The last equality follows the equation for the mean of a linear function in Chapter 5. However, because $E(X_i^2) = \mu^2 + \sigma^2$ and $E(\overline{X}^2) = \mu^2 + \sigma^2/n$, we have

$$E(S^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n}(\mu^2 + \sigma^2) - n(\mu^2 + \sigma^2/n)\right]$$

$$= \frac{1}{n-1}(n\mu^2 + n\sigma^2 - n\mu^2 - \sigma^2) = \sigma^2$$

Therefore, the sample variance S^2 is an unbiased estimator of the population variance σ^2.

Although S^2 is unbiased for σ^2, S is a biased estimator of σ. For large samples, the bias is very small. However, there are good reasons for using S as an estimator of σ in samples from normal distributions as we will see in the next three chapters when we discuss confidence intervals and hypothesis testing.

Sometimes there are several unbiased estimators of the sample population parameter. For example, suppose that we take a random sample of size $n = 10$ from a normal population and obtain the data $x_1 = 12.8$, $x_2 = 9.4$, $x_3 = 8.7$, $x_4 = 11.6$, $x_5 = 13.1$, $x_6 = 9.8$, $x_7 = 14.1$, $x_8 = 8.5$, $x_9 = 12.1$, $x_{10} = 10.3$. Now the sample mean is

$$\overline{x} = \frac{12.8 + 9.4 + 8.7 + 11.6 + 13.1 + 9.8 + 14.1 + 8.5 + 12.1 + 10.3}{10} = 11.04$$

the sample median is

$$\overline{x} = \frac{10.3 + 11.6}{2} = 10.95$$

and a 10% trimmed mean (obtained by discarding the smallest and largest 10% of the sample before averaging) is

$$\overline{x}_{tr(10)} = \frac{8.7 + 9.4 + 9.8 + 10.3 + 11.6 + 12.1 + 12.8 + 13.1}{8} = 10.98$$

We can show that all of these are unbiased estimates of μ. Because there is not a unique unbiased estimator, we cannot rely on the property of unbiasedness alone to select our estimator. We need a method to select among unbiased estimators. We suggest a method in the following section.

7.3.2 Variance of a Point Estimator

Suppose that $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are unbiased estimators of θ. This indicates that the distribution of each estimator is centered at the true value of zero. However, the variance of these distributions may be different. Figure 7.6 illustrates the situation. Because $\hat{\Theta}_1$ has a smaller variance than $\hat{\Theta}_2$, the estimator $\hat{\Theta}_1$ is more likely to produce an estimate close to the true value of θ. A logical principle of estimation when selecting among several unbiased estimators is to choose the estimator that has minimum variance.

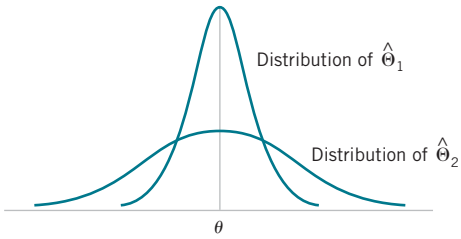


FIGURE 7.6

The sampling distributions of two unbiased estimators $\hat{\Theta}_1$ and $\hat{\Theta}_2$.

> **Minimum Variance Unbiased Estimator**
> If we consider all unbiased estimators of θ, the one with the smallest variance is called the **minimum variance unbiased estimator** (MVUE).

In a sense, the MVUE is most likely among all unbiased estimators to produce an estimate $\hat{\theta}$ that is close to the true value of θ. It has been possible to develop methodology to identify the MVUE in many practical situations. Although this methodology is beyond the scope of this book, we give one very important result concerning the normal distribution.

> If $X_1, X_2, \ldots, X_n$ is a random sample of size n from a normal distribution with mean μ and variance σ^2, the sample mean $\overline{X}$ is the MVUE for μ.

When we do not know whether an MVUE exists, we could still use a minimum variance principle to choose among competing estimators. Suppose, for example, we wish to estimate the mean of a population (not necessarily a *normal* population). We have a random sample of n observations $X_1, X_2, \ldots, X_n$, and we wish to compare two possible estimators for μ: the sample mean $\overline{X}$ and a single observation from the sample, say, X_i. Note that both $\overline{X}$ and X_i are unbiased estimators of μ; for the sample mean, we have $V(\overline{X}) = \sigma^2/n$ from Chapter 5 and the variance of any observation is $V(X_i) = \sigma^2$. Because $V(\overline{X}) < V(X_i)$ for sample sizes $n \geq 2$, we would conclude that the sample mean is a better estimator of μ than a single observation X_i.

7.3.3 Standard Error: Reporting a Point Estimate

When the numerical value or point estimate of a parameter is reported, it is usually desirable to give some idea of the precision of estimation. The measure of precision usually employed is the standard error of the estimator that has been used.

> **Standard Error of an Estimator**
> The **standard error** of an estimator $\hat{\Theta}$ is its standard deviation given by $\sigma_{\hat{\Theta}} = \sqrt{V(\hat{\Theta})}$. If the standard error involves unknown parameters that can be estimated, substitution of those values into $\sigma_{\hat{\Theta}}$ produces an **estimated standard error**, denoted by $\hat{\sigma}_{\hat{\Theta}}$.

Sometimes the estimated standard error is denoted by $s_{\hat{\Theta}}$ or $SE(\hat{\Theta})$.

Suppose that we are sampling from a normal distribution with mean μ and variance σ^2. Now the distribution of $\overline{X}$ is normal with mean μ and variance σ^2/n, so the **standard error** of $\overline{X}$ is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

If we did not know σ but substituted the sample standard deviation S into the preceding equation, the **estimated standard error** of $\overline{X}$ would be

$$SE(\overline{X}) = \hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n}}$$

When the estimator follows a normal distribution as in the preceding situation, we can be reasonably confident that the true value of the parameter lies within two standard errors of the estimate. Because many point estimators are normally distributed (or approximately so) for large n, this is a very useful result. Even when the point estimator is not normally distributed, we can state that so long as the estimator is unbiased, the estimate of the parameter will deviate from the true value by as much as four standard errors at most 6 percent of the time. Thus, a very conservative statement is that the true value of the parameter differs from the point estimate by at most four standard errors. See Chebyshev's inequality in the supplemental material on the Web site.

EXAMPLE 7.4 | Thermal Conductivity

An article in the *Journal of Heat Transfer* (Trans. ASME, Sec. C, 96, 1974, p. 59) described a new method of measuring the thermal conductivity of Armco iron. Using a temperature of 100°F and a power input of 550 watts, the following 10 measurements of thermal conductivity (in Btu/hr-ft-°F) were obtained:

41.60, 41.48, 42.34, 41.95, 41.86,
42.18, 41.72, 42.26, 41.81, 42.04

A point estimate of the mean thermal conductivity at 100°F and 550 watts is the sample mean or

$$\overline{x} = 41.924 \text{ Btu/hr-ft-°F}$$

The standard error of the sample mean is $\sigma_{\overline{X}} = \sigma/\sqrt{n}$, and because σ is unknown, we may replace it by the sample standard deviation $s = 0.284$ to obtain the estimated standard error of $\overline{X}$ as

$$SE(\overline{X}) = \hat{\sigma}_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{0.284}{\sqrt{10}} = 0.0898$$

Practical Interpretation: Notice that the standard error is about 0.2 percent of the sample mean, implying that we have obtained a relatively precise point estimate of thermal conductivity. If we can assume that thermal conductivity is normally distributed, two times the standard error is $2\hat{\sigma}_{\overline{X}} = 2(0.0898) = 0.1796$, and we are highly confident that the true mean thermal conductivity is within the interval 41.924 ± 0.1796 or between 41.744 and 42.104.

7.3.4 Bootstrap Standard Error

In some situations, the form of a point estimator is complicated, and standard statistical methods to find its standard error are difficult or impossible to apply. One example of these is S, the point estimator of the population standard deviation σ. Others occur with some of the standard probability distributions, such as the exponential and Weibull distributions. A relatively new **computer-intensive** technique, the **bootstrap**, can be used to solve this problem.

To explain how the bootstrap works, suppose that we have a random variable X with a known probability density function characterized by a parameter θ, say $f(x; \theta)$. Also assume that we have a random sample of data from this distribution, $x_1, x_2, \ldots, x_n$, and that the estimate

of θ based on this sample data is $\hat{\theta} = 4.5$. The bootstrap procedure would use the computer to generate **bootstrap samples** randomly from the probability distribution $f(x; \theta = 4.5)$ and calculate a **bootstrap estimate** $\hat{\theta}^B$. This process is repeated n_B times, resulting in:

$$\begin{aligned}
&\text{Bootstrap sample 1: } x_1^1, x_2^1 \ldots, x_n^1, \text{Bootstrap estimate } \hat{\theta}_1^B \\
&\text{Bootstrap sample 2: } x_1^2, x_2^2 \ldots, x_n^2, \text{Bootstrap estimate } \hat{\theta}_2^B \\
&\vdots \\
&\text{Bootstrap sample } n_B\text{: } x_1^{n_B}, x_2^{n_B} \ldots, x_n^{n_B}, \text{Bootstrap estimate } \hat{\theta}_{n_B}^B
\end{aligned}$$

Typically, the number of bootstrap samples is $n_B = 100$ or 200. The sample mean of the bootstrap estimates is

$$\bar{\theta}^B = \frac{1}{n_B} \sum_{i=1}^{n_B} \hat{\theta}_i^B$$

The **bootstrap standard error** of $\hat{\theta}$ is just the sample standard deviation of the bootstrap estimates $\hat{\theta}_i^B$ or

$$SE_B(\hat{\theta}) = \sqrt{\frac{1}{n_B - 1} \sum_{i=1}^{n_B} (\hat{\theta}_i^B - \bar{\theta}^B)^2} \tag{7.7}$$

Some authors use n_B in the denominator of Equation 7.7.

EXAMPLE 7.5 | Bootstrap Standard Error

A team of analytics specialists has been investigating the cycle time to process loan applications. The specialists' experience with the process informs them that cycle time is normally distributed with a mean of about 25 hours. A recent random sample of 10 applications gives the following (in hours):

24.1514, 27.4145, 20.4000, 22.5151, 28.5152, 28.5611, 21.2489, 20.9983, 24.9840, 22.6245

The sample standard deviation of these observations is $s = 3.11407$. We want to find a bootstrap standard error for the sample standard deviation. We use a computer program to generate $n_B = 200$ bootstrap samples from a normal distribution with a mean of 25 and a standard deviation of 3.11417. The first of these samples is:

25.4274, 24.2272, 24.8565, 24.3458, 18.4343, 23.3179, 23.0699, 25.2876, 27.1541, 27.2932

from which we calculate $s = 2.50635$. After all 200 bootstrap samples were generated, the average of the bootstrap estimates of the standard deviation was 3.03972, and the bootstrap estimate of the standard error was 0.5464. The standard error is fairly large because the sample size here ($n = 10$) is fairly small.

In some problem situations, the distribution of the random variable is not known. The bootstrap can still be used in these situations. The procedure is to treat the data sample as a population and draw bootstrap samples from it. So, for example, if we had a sample of 25 observations, we would draw n_B bootstrap samples by sampling with replacement from the original sample. Then we would proceed as in the preceding example to calculate the bootstrap estimate of the standard error for the statistic of interest.

7.3.5 Mean Squared Error of an Estimator

Sometimes it is necessary to use a biased estimator. In such cases, the mean squared error of the estimator can be important. The **mean squared error** of an estimator $\hat{\Theta}$ is the expected squared difference between $\hat{\Theta}$ and θ.

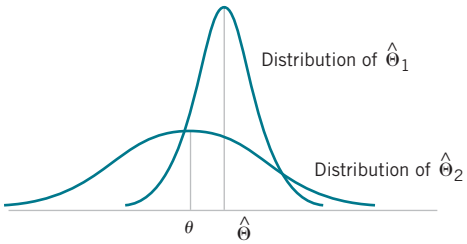


FIGURE 7.7

A biased estimator $\hat{\Theta}_1$ that has smaller variance than the unbiased estimator $\hat{\Theta}_2$.

Mean Squared Error of an Estimator

The **mean squared error of an estimator** $\hat{\Theta}$ of the parameter θ is defined as

$$\text{MSE}(\hat{\Theta}) = E(\hat{\Theta} - \theta)^2 \tag{7.8}$$

The mean squared error can be rewritten as follows:

$$\begin{aligned}\text{MSE}(\hat{\Theta}) &= E[\hat{\Theta} - E(\hat{\Theta})]^2 + [\theta - E(\hat{\Theta})]^2 \\ &= V(\hat{\Theta}) + (\text{bias})^2\end{aligned}$$

That is, the mean squared error of $\hat{\Theta}$ is equal to the variance of the estimator plus the squared bias. If $\hat{\Theta}$ is an unbiased estimator of θ, the mean squared error of $\hat{\Theta}$ is equal to the variance of $\hat{\Theta}$.

The mean squared error is an important criterion for comparing two estimators. Let $\hat{\Theta}_1$ and $\hat{\Theta}_2$ be two estimators of the parameter θ, and let MSE $(\hat{\Theta}_1)$ and MSE $(\hat{\Theta}_2)$ be the mean squared errors of $\hat{\Theta}_1$ and $\hat{\Theta}_2$. Then the **relative efficiency** of $\hat{\Theta}_2$ to $\hat{\Theta}_1$ is defined as

$$\frac{\text{MSE}(\hat{\Theta}_1)}{\text{MSE}(\hat{\Theta}_2)} \tag{7.9}$$

If this relative efficiency is less than 1, we would conclude that $\hat{\Theta}_1$ is a more efficient estimator of θ than $\hat{\Theta}_2$ in the sense that it has a smaller mean squared error.

Sometimes we find that biased estimators are preferable to unbiased estimators because they have smaller mean squared error. That is, we may be able to reduce the variance of the estimator considerably by introducing a relatively small amount of bias. As long as the reduction in variance is larger than the squared bias, an improved estimator from a mean squared error viewpoint will result. For example, Figure 7.7 is the probability distribution of a biased estimator $\hat{\Theta}_1$ that has a smaller variance than the unbiased estimator $\hat{\Theta}_2$. An estimate based on $\hat{\Theta}_1$ would more likely be close to the true value of θ than would an estimate based on $\hat{\Theta}_2$. Linear regression analysis (Chapters 11 and 12) is an example of an application area in which biased estimators are occasionally used.

An estimator $\hat{\Theta}$ that has a mean squared error that is less than or equal to the mean squared error of any other estimator, for all values of the parameter θ, is called an **optimal** estimator of θ. Optimal estimators rarely exist.

7.4 Methods of Point Estimation

The definitions of unbiasedness and other properties of estimators do not provide any guidance about how to obtain good estimators. In this section, we discuss methods for obtaining point estimators: the method of moments and the method of maximum likelihood. Maximum likelihood estimates are generally preferable to moment estimators because they have better efficiency properties. However, moment estimators are sometimes easier to compute. Both methods can produce unbiased point estimators. We also briefly discuss a Bayesian approach to parameter estimation in the WileyPLUS online material.

7.4.1 Method of Moments

The general idea behind the method of moments is to equate **population moments**, which are defined in terms of expected values, to the corresponding **sample moments**. The population moments will be functions of the unknown parameters. Then these equations are solved to yield estimators of the unknown parameters.

Moments

Let $X_1, X_2, \ldots, X_n$ be a random sample from the probability distribution $f(x)$ where $f(x)$ can be a discrete probability mass function or a continuous probability density function. The kth **population moment** (or **distribution moment**) is $E(X^k)$, $k = 1, 2, \ldots$. The corresponding kth **sample moment** is $(1/n)\sum_{i=1}^{n} X_i^k, k = 1, 2, \ldots$.

To illustrate, the first population moment is $E(X) = \mu$, and the first sample moment is $(1/n)\sum_{i=1}^{n} X_i = \overline{X}$. Thus, by equating the population and sample moments, we find that $\hat{\mu} = \overline{X}$. That is, the sample mean is the **moment estimator** of the population mean. In the general case, the population moments are functions of the unknown parameters of the distribution, say, $\theta_1, \theta_2, \ldots, \theta_m$.

Moment Estimators

Let $X_1, X_2, \ldots, X_n$ be a random sample from either a probability mass function or a probability density function with m unknown parameters $\theta_1, \theta_2, \ldots, \theta_m$. The **moment estimators** $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_m$ are found by equating the first m population moments to the first m sample moments and solving the resulting equations for the unknown parameters.

EXAMPLE 7.6 | Exponential Distribution Moment Estimator

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from an exponential distribution with parameter λ. Now there is only one parameter to estimate, so we must equate $E(X)$ to $\overline{X}$. For the exponential, $E(X) = 1/\lambda$. Therefore, $E(X) = \overline{X}$ results in $1/\lambda = \overline{X}$, so $\lambda = 1/\overline{X}$, is the moment estimator of λ.

As an example, suppose that the time to failure of an electronic module used in an automobile engine controller is tested at an elevated temperature to accelerate the failure mechanism. The time to failure is exponentially distributed. Eight units are randomly selected and tested, resulting in the following failure time (in hours): $x_1 = 11.96$, $x_2 = 5.03$, $x_3 = 67.40$, $x_4 = 16.07$, $x_5 = 31.50$, $x_6 = 7.73$, $x_7 = 11.10$, and $x_8 = 22.38$. Because $\overline{x} = 21.65$, the moment estimate of λ is $\hat{\lambda} = 1/\overline{x} = 1/21.65 = 0.0462$.

EXAMPLE 7.7 | Normal Distribution Moment Estimators

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with parameters μ and σ^2. For the normal distribution, $E(X) = \mu$ and $E(X^2) = \mu^2 + \sigma^2$. Equating $E(X)$ to $\overline{X}$ and $E(X^2)$ to $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ gives

$$\mu = \overline{X}, \qquad \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Solving these equations gives the moment estimators

$$\hat{\mu} = \overline{X}, \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2}{n} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n}$$

Practical Conclusion: Notice that the moment estimator of σ^2 is not an unbiased estimator.

EXAMPLE 7.8 | Gamma Distribution Moment Estimators

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a gamma distribution with parameters r and λ. For the gamma distribution, $E(X) = r/\lambda$ and $E(X^2) = r(r+1)/\lambda^2$. The moment estimators are found by solving

$$r/\lambda = \overline{X}, \quad r(r+1)/\lambda^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

The resulting estimators are

$$\hat{r} = \frac{\overline{X}^2}{(1/n)\sum_{i=1}^{n} X_i^2 - \overline{X}^2} \qquad \hat{\lambda} = \frac{\overline{X}}{(1/n)\sum_{i=1}^{n} X_i^2 - \overline{X}^2}$$

To illustrate, consider the time to failure data introduced following Example 7.6. For these data, $\overline{x} = 21.65$ and $\sum_{i=1}^{8} x_i^2 = 6639.40$, so the moment estimates are

$$\hat{r} = \frac{(21.65)^2}{(1/8)\,6645.43 - (21.65)^2} = 1.29,$$

$$\hat{\lambda} = \frac{21.65}{(1/8)\,6645.43 - (21.65)^2} = 0.0598$$

Interpretation: When $r = 1$, the gamma reduces to the exponential distribution. Because $\hat{r}$ slightly exceeds unity, it is quite possible that either the gamma or the exponential distribution would provide a reasonable model for the data.

7.4.2 Method of Maximum Likelihood

One of the best methods of obtaining a point estimator of a parameter is the method of maximum likelihood. This technique was developed in the 1920s by a famous British statistician, Sir R. A. Fisher. As the name implies, the estimator is the value of the parameter that maximizes the **likelihood function**.

Maximum Likelihood Estimator

Suppose that X is a random variable with probability distribution $f(x; \theta)$, where θ is a single unknown parameter. Let $x_1, x_2, \ldots, x_n$ be the observed values in a random sample of size n. Then the **likelihood function** of the sample is

$$L(\theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \cdots \cdot f(x_n; \theta) \tag{7.10}$$

Note that the likelihood function is now a function of only the unknown parameter θ. The **maximum likelihood estimator** (MLE) of θ is the value of θ that maximizes the likelihood function $L(\theta)$.

In the case of a discrete random variable, the interpretation of the likelihood function is simple. The likelihood function of the sample $L(\theta)$ is just the probability

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

That is, $L(\theta)$ is just the probability of obtaining the sample values $x_1, x_2, \ldots, x_n$. Therefore, in the discrete case, the maximum likelihood estimator is an estimator that maximizes the probability of occurrence of the sample values. Maximum likelihood estimators are generally preferable to moment estimators because they possess good efficiency properties.

EXAMPLE 7.9 | Bernoulli Distribution MLE

Let X be a Bernoulli random variable. The probability mass function is

$$f(x; p) = \begin{cases} p^x(1-p)^{1-x}, & x = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

where p is the parameter to be estimated. The likelihood function of a random sample of size n is

$$L(p) = p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \ldots p^{x_n}(1-p)^{1-x_n}$$
$$= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

We observe that if $\hat{p}$ maximizes $L(p)$, $\hat{p}$ also maximizes $\ln L(p)$. Therefore,

$$\ln L(p) = \left(\sum_{i=1}^{n} x_i\right) \ln p + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-p)$$

Now,

$$\frac{d \ln L(p)}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-p}$$

Equating this to zero and solving for p yields $\hat{p} = (1/n)\sum_{i=1}^{n} x_i$. Therefore, the maximum likelihood estimator of p is

$$\hat{P} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Suppose that this estimator were applied to the following situation: n items are selected at random from a production line, and each item is judged as either defective (in which case we set $x_i = 1$) or nondefective (in which case we set $x_i = 0$). Then $\sum_{i=1}^{n} x_i$ is the number of defective units in the sample, and $\hat{p}$ is the sample proportion defective. The parameter p is the population proportion defective, and it seems intuitively quite reasonable to use $\hat{p}$ as an estimate of p.

Although the interpretation of the likelihood function just given is confined to the discrete random variable case, the method of maximum likelihood can easily be extended to a continuous distribution. We now give two examples of maximum likelihood estimation for continuous distributions.

EXAMPLE 7.10 | Exponential Distribution MLE

Let X be exponentially distributed with parameter λ. The likelihood function of a random sample of size n, say, $X_1, X_2, \ldots, X_n$, is

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

The log likelihood is

$$\ln L(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i$$

Now

$$\frac{d \ln L(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

and upon equating this last result to zero, we obtain

$$\hat{\lambda} = n / \sum_{i=1}^{n} X_i = 1/\overline{X}$$

Conclusion: Thus, the maximum likelihood estimator of λ is the reciprocal of the sample mean. Notice that this is the same as the moment estimator.

It is easy to illustrate graphically just how the method of maximum likelihood works. Figure 7.8(a) plots the log of the likelihood function for the exponential parameter from Example 7.10, using the $n = 8$ observations on failure time given following Example 7.5. It is common for the log likelihood function to be negative. We found that the estimate of λ was $\hat{\lambda} = 0.0462$. From Example 7.10, we know that this is a maximum likelihood estimate. Figure 7.8(a) shows clearly that the log likelihood function is maximized at a value of λ that is approximately equal to 0.0462. Notice that the log likelihood function is relatively flat in the region of the maximum. This implies that the parameter is not estimated very precisely. If the parameter were estimated precisely, the log likelihood function would be very peaked at the maximum value. The sample size here is relatively small, and this has led to the imprecision in estimation. This is illustrated in Figure 7.8(b) where we have plotted the difference in log likelihoods for the maximum

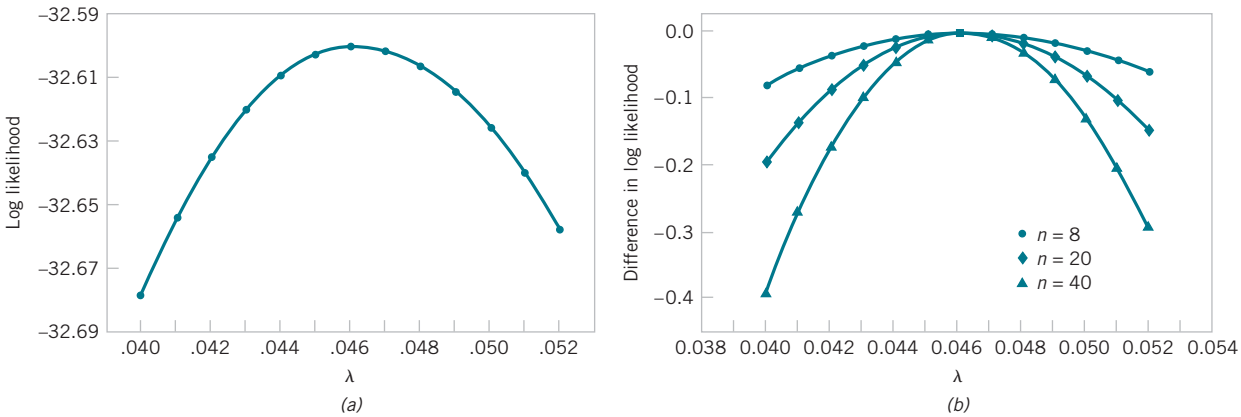


FIGURE 7.8

Log likelihood for the exponential distribution, using the failure time data. (a) Log likelihood with $n = 8$ (original data). (b) Log likelihood if $n = 8$, 20, and 40.

value, assuming that the sample sizes were $n = 8$, 20, and 40 but that the sample average time to failure remained constant at $\bar{x} = 21.65$. Notice how much steeper the log likelihood is for $n = 20$ in comparison to $n = 8$, and for $n = 40$ in comparison to both smaller sample sizes.

The method of maximum likelihood can be used in situations that have several unknown parameters, say, $\theta_1, \theta_2, \ldots, \theta_k$, to estimate. In such cases, the likelihood function is a function of the k unknown parameters $\theta_1, \theta_2, \ldots, \theta_k$, and the maximum likelihood estimators $\{\hat{\Theta}_i\}$ would be found by equating the k partial derivatives $\partial L(\theta_1, \theta_2, \ldots, \theta_k)/\partial\theta_i, i = 1, 2, \ldots, k$ to zero and solving the resulting system of equations.

EXAMPLE 7.11 | Normal Distribution MLEs for μ and σ^2

Let X be normally distributed with mean μ and variance σ^2 where both μ and σ^2 are unknown. The likelihood function for a random sample of size n is

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/(2\sigma^2)} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

and

$$\ln L(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

Now

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial(\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

The solutions to these equations yield the maximum likelihood estimators

$$\hat{\mu} = \bar{X} \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Conclusion: Once again, the maximum likelihood estimators are equal to the moment estimators.

Properties of the Maximum Likelihood Estimator As noted previously, the method of maximum likelihood is often the estimation method that we prefer because it produces estimators with good statistical properties. We summarize these properties as follows.

Properties of a Maximum Likelihood Estimator

Under very general and not restrictive conditions when the sample size n is large and if $\hat{\Theta}$ is the maximum likelihood estimator of the parameter θ,

(1) $\hat{\Theta}$ is an approximately unbiased estimator for θ ($E(\hat{\Theta}) = \theta$),
(2) The variance of $\hat{\Theta}$ is nearly as small as the variance that could be obtained with any other estimator.
(3) $\hat{\Theta}$ has an approximate normal distribution.

Properties 1 and 2 essentially state that the maximum likelihood estimator is approximately an MVUE. This is a very desirable result and, coupled with the facts that it is fairly easy to obtain in many situations and has an asymptotic normal distribution (the "asymptotic" means "when n is large"), explains why the maximum likelihood estimation technique is widely used. To use maximum likelihood estimation, remember that the distribution of the population must be either known or assumed.

To illustrate the "large-sample" or asymptotic nature of these properties, consider the maximum likelihood estimator for σ^2, the variance of the normal distribution, in Example 7.11. It is easy to show that

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$$

The bias is

$$E(\hat{\sigma}^2) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = \frac{-\sigma^2}{n}$$

Because the bias is negative, $\hat{\sigma}^2$ tends to underestimate the true variance σ^2. Note that the bias approaches zero as n increases. Therefore, $\hat{\sigma}^2$ is an asymptotically unbiased estimator for σ^2.

We now give another very important and useful property of maximum likelihood estimators.

Invariance Property

Let $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k$ be the maximum likelihood estimators of the parameters $\theta_1, \theta_2, \ldots, \theta_k$. Then the maximum likelihood estimator of any function $h(\theta_1, \theta_2, \ldots, \theta_k)$ of these parameters is the same function $h(\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k)$ of the estimators $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k$.

EXAMPLE 7.12

In the normal distribution case, the maximum likelihood estimators of μ and σ^2 were $\hat{\mu} = \overline{X}$ and $\hat{\sigma}^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/n$, respectively. To obtain the maximum likelihood estimator of the function $h(\mu, \sigma^2) = \sqrt{\sigma^2} = \sigma$, substitute the estimators $\hat{\mu}$ and $\hat{\sigma}^2$ into the function h, which yields

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right]^{1/2}$$

Conclusion: The maximum likelihood estimator of the standard deviation σ is *not* the sample standard deviation S.

Complications in Using Maximum Likelihood Estimation Although the method of maximum likelihood is an excellent technique, sometimes complications arise in its use. For example, it is not always easy to maximize the likelihood function because the equation(s) obtained from $dL(\theta)/d\theta = 0$ may be difficult to solve. Furthermore, it may not always be possible

to use calculus methods directly to determine the maximum of $L(\theta)$. This is illustrated in the following example.

EXAMPLE 7.13 | Uniform Distribution MLE

Let X be uniformly distributed on the interval 0 to a. Because the density function is $f(x) = 1/a$ for $0 \le x \le a$ and zero otherwise, the likelihood function of a random sample of size n is

$$L(a) = \prod_{i=1}^{n} \frac{1}{a} = \frac{1}{a^n}$$

for

$$0 \le x_1 \le a, 0 \le x_2 \le a, \ldots, 0 \le x_n \le a$$

Note that the slope of this function is not zero anywhere. That is, as long as $\max(x_i) \le a$, the likelihood is $1/a^n$, which is positive, but when $a < \max(x_i)$, the likelihood goes to zero as illustrated in Figure 7.9. Therefore, calculus methods cannot be used directly because the maximum value of the likelihood function occurs at a point of discontinuity. However, because $d/da(a^{-n}) = -n/a^{n+1}$ is less than zero for all values of $a > 0$, a^{-n} is a decreasing function of a. This implies that the maximum of the likelihood function $L(a)$ occurs at the lower boundary point. The figure clearly shows that we could maximize $L(a)$ by setting $\hat{a}$ equal to the smallest value that it could logically take on, which is $\max(x_i)$. Clearly, a cannot be smaller than the largest sample observation, so setting $\hat{a}$ equal to the largest sample value is reasonable.

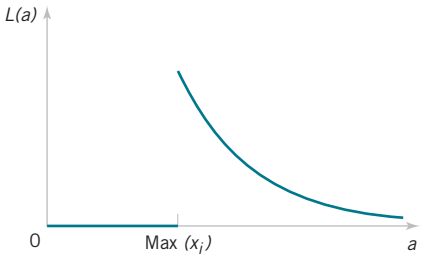


FIGURE 7.9

The likelihood function for the uniform distribution in Example 7.13.

7.4.3 Bayesian Estimation of Parameters

This book uses methods of statistical inference based on the information in the sample data. In effect, these methods interpret probabilities as relative frequencies. Sometimes we call probabilities that are interpreted in this manner **objective probabilities**. Another approach to statistical inference, called the **Bayesian** approach, combines sample information with other information that may be available prior to collecting the sample. In this section, we briefly illustrate how this approach may be used in parameter estimation.

Suppose that the random variable X has a probability distribution that is a function of one parameter θ. We write this probability distribution as $f(x|\theta)$. This notation implies that the exact form of the distribution of X is conditional on the value assigned to θ. The classical approach to estimation would consist of taking a random sample of size n from this distribution and then substituting the sample values x_i into the estimator for θ. This estimator could have been developed using the maximum likelihood approach, for example.

Suppose that we have some additional information about θ and that we can summarize that information in the form of a probability distribution for θ, say, $f(\theta)$. This probability distribution is often called the **prior distribution** for θ, and suppose that the mean of the prior is μ_0 and the variance is σ_0^2. This is a very novel concept insofar as the rest of this book is concerned because we are now viewing the parameter θ as a random variable. The probabilities associated with the prior distribution are often called **subjective probabilities** because they usually reflect the analyst's degree of belief regarding the true value of θ. The Bayesian approach to estimation uses the prior distribution for θ, $f(\theta)$, and the joint probability distribution of the sample, say, $f(x_1, x_2, \ldots, x_n|\theta)$, to find a **posterior distribution** for θ, say, $f(\theta|x_1, x_2, \ldots, x_n)$. This posterior distribution contains information from both the sample and the prior distribution for θ. In a sense, it expresses our degree of belief regarding the true value of θ after observing the sample data. It is easy conceptually to find the posterior distribution. The joint probability distribution of the sample $X_1, X_2, \ldots, X_n$ and the parameter θ (remember that θ is a random variable) is

$$f(x_1, x_2, \ldots, x_n, \theta) = f(x_1, x_2, \ldots, x_n|\theta) f(\theta)$$

and the marginal distribution of $X_1, X_2, \ldots, X_n$ is

$$f(x_1, x_2, \ldots, x_n) = \begin{cases} \sum_{\theta} f(x_1, x_2, \ldots, x_n, \theta), & \theta \text{ discrete} \\ \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n, \theta)\, d\theta, & \theta \text{ continuous} \end{cases}$$

Therefore, the desired distribution is

$$f(\theta | x_1, x_2, \ldots, x_n) = \frac{f(x_1, x_2, \ldots, x_n, \theta)}{f(x_1, x_2, \ldots, x_n)}$$

We define the **Bayes estimator** of θ as the value θ that corresponds to the mean of the posterior distribution $f(\theta | x_1, x_2, \ldots, x_n)$.

Sometimes the mean of the posterior distribution of θ can be determined easily. As a function of θ, $f(\theta | x_1, \ldots, x_n)$ is a probability density function and $x_1, \ldots, x_n$ are just constants. Because θ enters into $f(\theta | x_1, \ldots, x_n)$ only through $f(x_1, \ldots, x_n, \theta)$ if $f(x_1, \ldots, x_n, \theta)$ is recognized as a well-known probability function, the posterior mean of θ can be deduced from the well-known distribution without integration or even calculation of $f(x_1, \ldots, x_n)$.

EXAMPLE 7.14 | Bayes Estimator for the Mean of a Normal Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from the normal distribution with mean μ and variance σ^2 where μ is unknown and σ^2 is known. Assume that the prior distribution for μ is normal with mean μ_0 and variance σ_0^2; that is,

$$f(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-(\mu-\mu_0)^2/(2\sigma_0^2)} = \frac{1}{\sqrt{2\pi}\sigma_0^2} e^{-(\mu^2 - 2\mu_0 + \mu_0^2)/(2\sigma_0^2)}$$

The joint probability distribution of the sample is

$$f(x_1, x_2, \ldots, x_n | \mu) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)\sum_{i=1}^{n}(x_i-\mu)^2}$$
$$= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)(\sum x_i^2 - 2\mu\sum x_i + n\mu^2)}$$

Thus, the joint probability distribution of the sample and μ is

$$f(x_1, x_2, \ldots, x_n, \mu) = \frac{1}{(2\pi\sigma^2)^{n/2}\sqrt{2\pi}\sigma_0} \times e^{-(1/2)\left[(1/\sigma_0^2 + n/\sigma^2)\mu^2 - \left(2\mu_0/\sigma_0^2 + 2\sum x_i/\sigma^2\right)\mu + \sum x_i^2/\sigma^2 + \mu_0^2/\sigma_0^2\right]}$$
$$= e^{-(1/2)\left[\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right)\mu^2 - 2\left(\frac{\mu_0}{\sigma_0^2} + \frac{\bar{x}}{\sigma^2/n}\right)\mu\right]} h_1(x_1, \ldots, x_n, \sigma^2, \mu_0, \sigma_0^2)$$

Upon completing the square in the exponent,

$$f(x_1, x_2, \ldots, x_n, \mu) = e^{-(1/2)\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right)\left[\mu^2 - \left(\frac{(\sigma^2/n)\mu_0}{\sigma_0^2 + \sigma^2/n} + \frac{\bar{x}\sigma_0^2}{\sigma_0^2 + \sigma^2/n}\right)\right]^2} h_2(x_1, \ldots, x_n, \sigma^2, \mu_0, \sigma_0^2)$$

where $h_i(x_1, \ldots, x_n, \sigma^2, \mu_0, \sigma_0^2)$ is a function of the observed values and the parameters σ^2, μ_0, and σ_0^2.

Now, because $f(x_1, \ldots, x_n)$ does not depend on μ,

$$f(\mu | x_1, \ldots, x_n) = e^{-(1/2)\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right)\left[\mu^2 - \left(\frac{(\sigma^2/n)\mu_0 + \sigma_0^2\bar{x}}{\sigma_0^2 + \sigma^2/n}\right)\right]} h_3(x_1, \ldots, x_n, \sigma^2, \mu_0, \sigma_0^2)$$

This is recognized as a normal probability density function with posterior mean

$$\frac{(\sigma^2/n)\mu_0 + \sigma_0^2\bar{x}}{\sigma_0^2 + \sigma^2/n}$$

and posterior variance

$$\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right)^{-1} = \frac{\sigma_0^2(\sigma^2/n)}{\sigma_0^2 + \sigma^2/n}$$

Consequently, the Bayes estimate of μ is a weighted average of μ_0 and $\bar{x}$. For purposes of comparison, note that the maximum likelihood estimate of μ is $\hat{\mu} = \bar{x}$.

To illustrate, suppose that we have a sample of size $n = 10$ from a normal distribution with unknown mean μ and variance $\sigma^2 = 4$. Assume that the prior distribution for μ is normal with mean $\mu_0 = 0$ and variance $\sigma_0^2 = 1$. If the sample mean is 0.75, the Bayes estimate of μ is

$$\frac{(4/10)0 + 1(0.75)}{1 + (4/10)} = \frac{0.75}{1.4} = 0.536$$

Conclusion: Note that the maximum likelihood estimate of μ is $\bar{x} = 0.75$. The Bayes estimate is between the maximum likelihood estimate and the prior mean.

A relationship exists between the Bayes estimator for a parameter and the maximum likelihood estimator of the same parameter. For large sample sizes, the two are nearly equivalent. In general, the difference between the two estimators is small compared to $1/\sqrt{n}$. In practical problems, a moderate sample size will produce approximately the same estimate by either the Bayes or maximum likelihood method if the sample results are consistent with the assumed prior information. If the sample results are inconsistent with the prior assumptions, the Bayes estimate may differ considerably from the maximum likelihood estimate. In these circumstances, if the sample results are accepted as being correct, the prior information must be incorrect. The maximum likelihood estimate would then be the better estimate to use.

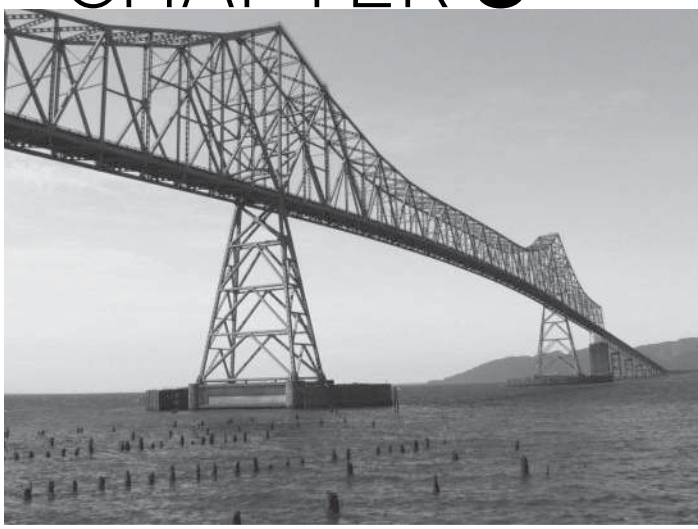
If the sample results are very different from the prior information, the Bayes estimator will always tend to produce an estimate that is between the maximum likelihood estimate and the prior assumptions. If there is more inconsistency between the prior information and the sample, there will be more difference between the two estimates.

Important Terms and Concepts

Bayes estimator
Bias in parameter estimation
Bootstrap method
Central limit theorem
Estimator versus estimate
Likelihood function
Maximum likelihood estimator
Mean squared error of an estimator
Minimum variance unbiased estimator
Moment estimator
Normal distribution as the sampling distribution of the difference in two sample means
Normal distribution as the sampling distribution of a sample mean
Parameter estimation
Point estimator
Population or distribution moments
Posterior distribution
Prior distribution
Sample moments
Sampling distribution
Standard error and estimated standard error of an estimator
Statistic
Statistical inference
Unbiased estimator

CHAPTER **8**

Statistical Intervals for a Single Sample



© jkullander / iStockphoto

LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Construct confidence intervals on the mean of a normal distribution, using either the normal distribution or the *t* distribution method
2. Construct confidence intervals on the variance and standard deviation of a normal distribution
3. Construct confidence intervals on a population proportion
4. Use a general method for constructing an approximate confidence interval on a parameter
5. Construct a prediction interval for a future observation
6. Construct a tolerance interval for a normal distribution
7. Explain the three types of interval estimates: confidence intervals, prediction intervals, and tolerance intervals

CHAPTER OUTLINE

8.1 Confidence Interval on the Mean of a Normal Distribution, Variance Known
- 8.1.1 Development of the Confidence Interval and Its Basic Properties
- 8.1.2 Choice of Sample Size
- 8.1.3 One-Sided Confidence Bounds
- 8.1.4 General Method to Derive a Confidence Interval
- 8.1.5 Large-Sample Confidence Interval for μ

8.2 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown
- 8.2.1 *t* Distribution
- 8.2.2 *t* Confidence Interval on μ

8.3 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

8.4 Large-Sample Confidence Interval for a Population Proportion

8.5 Guidelines for Constructing Confidence Intervals

8.6 Bootstrap Confidence Interval

8.7 Tolerance and Prediction Intervals
- 8.7.1 Prediction Interval for a Future Observation
- 8.7.2 Tolerance Interval for a Normal Distribution

Introduction

Engineers are often involved in estimating parameters. For example, there is an ASTM Standard E23 that defines a technique called the *Charpy V-notch method for notched bar impact testing* of metallic materials. The impact energy is often used to determine whether the material experiences a ductile-to-brittle transition as the temperature decreases. Suppose that we have tested a sample of 10 specimens of a particular material with this procedure. We know that we can use the sample average $\overline{X}$ to estimate the true mean impact energy μ. However, we also know that the true mean impact energy is unlikely to be exactly equal to your estimate. Reporting the results of your test as a single number is unappealing because nothing inherent in $\overline{X}$ provides any information about how close it is to the true value of the mean μ. Our estimate could be very close, or it could be considerably far from the true mean. A way to avoid this is to report the estimate in terms of a range of plausible values called a **confidence interval**. A confidence interval always specifies a confidence level, usually 90%, 95%, or 99%, which is a measure of the reliability of the procedure. So if a 95% confidence interval on the impact energy based on the data from our 10 specimens has a lower limit of 63.84*J* and an upper limit of 65.08*J*, then we can say that at the 95% level of confidence any value of *mean* impact energy between 63.84*J* and 65.08*J* is a plausible value. By *reliability*, we mean that if we repeated this experiment over and over again, 95% of all samples would produce a confidence interval that contains the true mean impact energy, and only 5% of the time would the interval be in error. In this chapter, you learn how to construct confidence intervals and other useful types of statistical intervals for many important types of problem situations.

In the previous chapter, we illustrated how a point estimate of a parameter can be estimated from sample data. However, it is important to understand how good the estimate obtained is. For example, suppose that we estimate the mean viscosity of a chemical product to be $\hat{\mu} = \bar{x} = 1000$. Now because of sampling variability, it is almost never the case that the true mean μ is exactly equal to the estimate $\bar{x}$. The point estimate says nothing about how close $\hat{\mu}$ is to μ. Is the process mean likely to be between 900 and 1100? Or is it likely to be between 990 and 1010? The answer to these questions affects our decisions regarding this process. Bounds that represent an interval of plausible values for a parameter are examples of an interval estimate. Surprisingly, it is easy to determine such intervals in many cases, and the same data that provided the point estimate are typically used.

An interval estimate for a population parameter is called a **confidence interval**. Information about the precision of estimation is conveyed by the length of the interval. A short interval implies precise estimation. We cannot be certain that the interval contains the true, unknown population

parameter—we use only a sample from the full population to compute the point estimate and the interval. However, the confidence interval is constructed so that we have high confidence that it does contain the unknown population parameter. Confidence intervals are widely used in engineering and the sciences.

A **tolerance interval** is another important type of interval estimate. For example, the chemical product viscosity data might be assumed to be normally distributed. We might like to calculate limits that bound 95% of the viscosity values. For a normal distribution, we know that 95% of the distribution is in the interval

$$\mu - 1.96\sigma, \mu - 19.6\sigma$$

However, this is not a useful tolerance interval because the parameters μ and σ are unknown. Point estimates such as $\bar{x}$ and s can be used in the preceding equation for μ and σ. However, we need to account for the potential error in each point estimate to form a tolerance interval for the distribution. The result is an interval of the form

$$\bar{x} - ks,\ \bar{x} + ks$$

where k is an appropriate constant (that is larger than 1.96 to account for the estimation error). As in the case of a confidence interval, it is not certain that the tolerance interval bounds 95% of the distribution, but the interval is constructed so that we have high confidence that it does. Tolerance intervals are widely used and, as we will subsequently see, they are easy to calculate for normal distributions.

Confidence and tolerance intervals bound unknown elements of a distribution. In this chapter, you learn to appreciate the value of these intervals. A **prediction interval** provides bounds on one (or more) *future observations* from the population. For example, a prediction interval could be used to bound a single, new measurement of viscosity—another useful interval. With a large sample size, the prediction interval for normally distributed data tends to the tolerance interval, but for more modest sample sizes, the prediction and tolerance intervals are different.

Keep the purpose of the three types of interval estimates clear:

- A confidence interval bounds population or distribution parameters (such as the mean viscosity).
- A tolerance interval bounds a selected proportion of a distribution.
- A prediction interval bounds future observations from the population or distribution.

Our experience has been that it is easy to confuse the three types of intervals. For example, a confidence interval is often reported when the problem situation calls for a prediction interval. Confidence intervals are covered in this chapter, while tolerance and prediction intervals are presented in the online material.

8.1 Confidence Interval on the Mean of a Normal Distribution, Variance Known

The basic ideas of a confidence interval (CI) are most easily understood by initially considering a simple situation. Suppose that we have a normal population with unknown mean μ and known variance σ^2. This is a somewhat unrealistic scenario because typically both the mean and variance are unknown. However, in subsequent sections, we present confidence intervals for more general situations.

8.1.1 Development of the Confidence Interval and Its Basic Properties

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with unknown mean μ and known variance σ^2. From the results of Chapter 5, we know that the sample mean $\overline{X}$ is

normally distributed with mean μ and variance σ^2/n. We may *standardize* $\overline{X}$ by subtracting the mean and dividing by the standard deviation, which results in the variable

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{8.1}$$

The random variable Z has a standard normal distribution.

A **confidence interval** estimate for μ is an interval of the form $l \le \mu \le u$, where the end-points l and u are computed from the sample data. Because different samples will produce different values of l and u, these end-points are values of random variables L and U, respectively. Suppose that we can determine values of L and U such that the following probability statement is true:

$$P\{L \le \mu \le U\} = 1 - \alpha \tag{8.2}$$

where $0 \le \alpha \le 1$. There is a probability of $1 - \alpha$ of selecting a sample for which the CI will contain the true value of μ. Once we have selected the sample, so that $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, and computed l and u, the resulting confidence interval for μ is

$$l \le \mu \le u \tag{8.3}$$

The end-points or bounds l and u are called the **lower**- and **upper-confidence limits (bounds)**, respectively, and $1 - \alpha$ is called the **confidence coefficient**.

In our problem situation, because $Z = \left(\overline{X} - \mu\right)/\left(\sigma/\sqrt{n}\right)$ has a standard normal distribution, we may write

$$P\left\{-z_{\alpha/2} \le \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right\} = 1 - \alpha$$

Now manipulate the quantities inside the brackets by (1) multiplying through by $\sigma/\sqrt{n}$, (2) subtracting $\overline{X}$ from each term, and (3) multiplying through by -1. This results in

$$P\left\{\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha \tag{8.4}$$

This is a *random interval* because the end-points $\overline{X} \pm Z_{\alpha/2}\sigma/\sqrt{n}$ involve the random variable $\overline{X}$. From consideration of Equation 8.4, the lower and upper end-points or limits of the inequalities in Equation 8.4 are the lower- and upper-confidence limits L and U, respectively. This leads to the following definition.

Confidence Interval on the Mean, Variance Known

If $\bar{x}$ is the sample mean of a random sample of size n from a normal population with known variance σ^2, a $100(1 - \alpha)\%$ confidence interval on μ is given by

$$\bar{x} - z_{\alpha/2}\sigma/\sqrt{n} \le \mu \le \bar{x} + z_{\alpha/2}\sigma/\sqrt{n} \tag{8.5}$$

where $z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution.

The development of this CI assumed that we are sampling from a normal population. The CI is quite robust to this assumption. That is, moderate departures from normality are of no serious concern. From a practical viewpoint, this implies that an *advertised* 95% CI might have actual confidence of 93% or 94%.

EXAMPLE 8.1 | Metallic Material Transition

ASTM Standard E23 defines standard test methods for notched bar impact testing of metallic materials. The Charpy V-notch (CVN) technique measures impact energy and is often used to determine whether or not a material experiences a ductile-to-brittle transition with decreasing temperature. Ten measurements of impact energy (J) on specimens of A238 steel cut at 60°C are as follows: 64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, and 64.3. Assume that impact energy is normally distributed with $\sigma = 1\ J$. We want to find a 95% CI for μ, the mean impact energy. The required quantities are $z_{\alpha/2} = z_{0.025} = 1.96$, $n = 10$, $\sigma = 1$, and $\bar{x} = 64.46$.

The resulting 95% CI is found from Equation 8.5 as follows:

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$64.46 - 1.96\frac{1}{\sqrt{10}} \le \mu \le 64.46 + 1.96\frac{1}{\sqrt{10}}$$

$$63.84 \le \mu \le 65.08$$

Practical Interpretation: Based on the sample data, a range of highly plausible values for mean impact energy for A238 steel at 60°C is $63.84\ J \le \mu \le 65.08\ J$.

Interpreting a Confidence Interval How does one interpret a confidence interval? In the impact energy estimation problem in Example 8.1, the 95% CI is $63.84 \le \mu \le 65.08$, so it is tempting to conclude that μ is within this interval with probability 0.95. However, with a little reflection, it is easy to see that this cannot be correct; the true value of μ is unknown, and the statement $63.84 \le \mu \le 65.08$ is either correct (true with probability 1) or incorrect (false with probability 1). The correct interpretation lies in the realization that a CI is a *random interval* because in the probability statement defining the end-points of the interval (Equation 8.2), L and U are random variables. Consequently, the correct interpretation of a $100(1-\alpha)\%$ CI depends on the relative frequency view of probability. Specifically, if an infinite number of random samples are collected and a $100(1-\alpha)\%$ confidence interval for μ is computed from each sample, $100(1-\alpha)\%$ of these intervals will contain the true value of μ.

The situation is illustrated in Figure 8.1, which shows several $100(1-\alpha)\%$ confidence intervals for the mean μ of a normal distribution. The dots at the center of the intervals indicate the point estimate of μ (that is, $\bar{x}$). Notice that one of the intervals fails to contain the true value of μ. If this were a 95% confidence interval, in the long run only 5% of the intervals would fail to contain μ.

Now in practice, we obtain only one random sample and calculate one confidence interval. Because this interval either will or will not contain the true value of μ, it is not reasonable to attach a probability level to this specific event. The appropriate statement is that the observed interval $[l, u]$ brackets the true value of μ with *confidence* $100(1-\alpha)$. This statement has a frequency interpretation; that is, we do not know whether the statement is true for this specific sample, but the *method* used to obtain the interval $[l, u]$ yields correct statements $100(1-\alpha)\%$ of the time.

Confidence Level and Precision of Estimation Notice that in Example 8.1, our choice of the 95% level of confidence was essentially arbitrary. What would have happened if we had chosen a higher level of confidence, say, 99%? In fact, is it not reasonable that we would want the

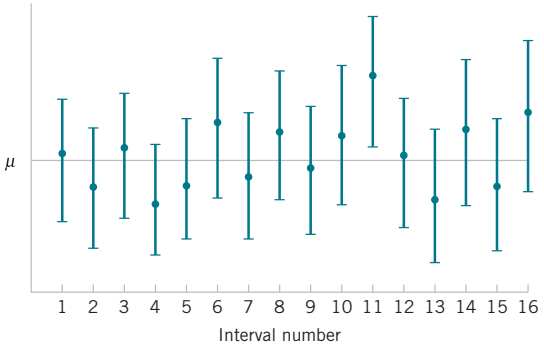


FIGURE 8.1
Repeated construction of a confidence interval for μ.

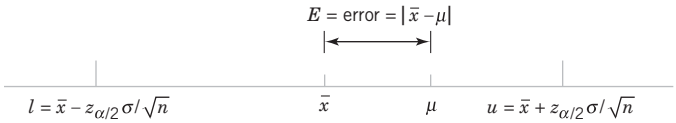


FIGURE 8.2
Error in estimating μ with $\bar{x}$.

higher level of confidence? At $\alpha = 0.01$, we find $z_{\alpha/2} = z_{0.01/2} = z_{0.005} = 2.58$, while for $\alpha = 0.05$, $z_{0.025} = 1.96$. Thus, the *length* of the 95% confidence interval is

$$2\left(1.96\sigma/\sqrt{n}\right) = 3.92\sigma/\sqrt{n}$$

whereas the length of the 99% CI is

$$2\left(2.58\sigma/\sqrt{n}\right) = 5.16\sigma/\sqrt{n}$$

Thus, the 99% CI is longer than the 95% CI. This is why we have a higher level of confidence in the 99% confidence interval. Generally, for a fixed sample size n and standard deviation σ, the higher the confidence level, the longer the resulting CI.

The length of a confidence interval is a measure of the *precision* of estimation. Many authors define the half-length of the CI (in our case $z_{\alpha/2}\sigma/\sqrt{n}$) as the bound on the error in estimation of the parameter. From the preceeding discussion, we see that precision is inversely related to the confidence level. It is desirable to obtain a confidence interval that is short enough for decision-making purposes and that also has adequate confidence. One way to achieve this is by choosing the sample size n to be large enough to give a CI of specified length or precision with prescribed confidence.

8.1.2 Choice of Sample Size

The precision of the confidence interval in Equation 8.5 is $2z_{\alpha/2}\sigma/\sqrt{n}$. This means that in using $\bar{x}$ to estimate μ, the error $E = |\bar{x} - \mu|$ is less than or equal to $z_{\alpha/2}\sigma/\sqrt{n}$ with confidence $100(1 - \alpha)$. This is shown graphically in Figure 8.2. In situations whose sample size can be controlled, we can choose n so that we are $100(1 - \alpha)\%$ confident that the error in estimating μ is less than a specified bound on the error E. The appropriate sample size is found by choosing n such that $z_{\alpha/2}\sigma/\sqrt{n} = E$. Solving this equation gives the following formula for n.

Sample Size for Specified Error on the Mean, Variance Known

If $\bar{x}$ is used as an estimate of μ, we can be $100(1 - \alpha)\%$ confident that the error $|\bar{x} - \mu|$ will not exceed a specified amount E when the sample size is

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 \tag{8.6}$$

If the right-hand side of Equation 8.6 is not an integer, it must be rounded up. This will ensure that the level of confidence does not fall below $100(1 - \alpha)\%$. Notice that $2E$ is the length of the resulting confidence interval.

EXAMPLE 8.2 | Metallic Material Transition

To illustrate the use of this procedure, consider the CVN test described in Example 8.1 and suppose that we want to determine how many specimens must be tested to ensure that the 95% CI on μ for A238 steel cut at 60°C has a length of at most 1.0 *J*. Because the bound on error in estimation E is one-half of the length of the CI, to determine n, we use Equation 8.6 with $E = 0.5$, $\sigma = 1$, and $z_{\alpha/2} = 1.96$. The required sample size is

$$n = \left(\frac{z_{\alpha/2}\,\sigma}{E}\right)^2 = \left[\frac{(1.96)1}{0.5}\right]^2 = 15.37$$

and because n must be an integer, the required sample size is $n = 16$.

Notice the general relationship between sample size, desired length of the confidence interval $2E$, confidence level $100(1-\alpha)$, and standard deviation σ:

- As the desired length of the interval $2E$ decreases, the required sample size n increases for a fixed value of σ and specified confidence.
- As σ increases, the required sample size n increases for a fixed desired length $2E$ and specified confidence.
- As the level of confidence increases, the required sample size n increases for fixed desired length $2E$ and standard deviation σ.

8.1.3 One-Sided Confidence Bounds

The confidence interval in Equation 8.5 gives both a lower confidence bound and an upper confidence bound for μ. Thus, it provides a two-sided CI. It is also possible to obtain one-sided confidence bounds for μ by setting either the lower bound $l = -\infty$ or the upper bound $u = \infty$ and replacing $z_{\alpha/2}$ by z_α.

One-Sided Confidence Bounds on the Mean, Variance Known

A $100(1-\alpha)\%$ **upper-confidence bound** for μ is

$$\mu \le \bar{x} + z_\alpha \sigma/\sqrt{n} \tag{8.7}$$

and a $100(1-\alpha)\%$ **lower-confidence bound** for μ is

$$\bar{x} - z_\alpha \sigma/\sqrt{n} \le \mu \tag{8.8}$$

EXAMPLE 8.3 | One-Sided Confidence Bound

The same data for impact testing from Example 8.1 are used to construct a lower, one-sided 95% confidence interval for the mean impact energy. Recall that $\bar{x} = 64.46, \sigma = 1J$, and $n = 10$. The interval is

$$\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} \le \mu$$

$$64.46 - 1.64 \frac{1}{\sqrt{10}} \le \mu$$

$$63.94 \le \mu$$

Practical Interpretation: The lower limit for the two-sided interval in Example 8.1 was 63.84. Because $z_\alpha < z_{\alpha/2}$, the lower limit of a one-sided interval is always greater than the lower limit of a two-sided interval of equal confidence. The one-sided interval does not bound μ from above so that it still achieves 95% confidence with a slightly larger lower limit. If our interest is only in the lower limit for μ, then the one-sided interval is preferred because it provides equal confidence with a greater limit. Similarly, a one-sided upper limit is always less than a two-sided upper limit of equal confidence.

8.1.4 General Method to Derive a Confidence Interval

It is easy to give a general method for finding a confidence interval for an unknown parameter θ. Let $X_1, X_2, \ldots, X_n$ be a random sample of n observations. Suppose that we can find a statistic $g(X_1, X_2, \ldots, X_n; \theta)$ with the following properties:

1. $g(X_1, X_2, \ldots, X_n; \theta)$ depends on both the sample and θ.
2. The probability distribution of $g(X_1, X_2, \ldots, X_n; \theta)$ does not depend on θ or any other unknown parameter.

In the case considered in this section, the parameter was $\theta = \mu$. The random variable $g(X_1, X_2, \ldots, X_n; \mu) = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ satisfies both conditions; the random variable depends on the sample and on μ, and it has a standard normal distribution because σ is known. Now we must find constants C_L and C_U so that

$$P[C_L \le g(X_1, X_2, \ldots, X_n;\ \theta) \le C_U] = 1 - \alpha \tag{8.9}$$

Because of property 2, C_L and C_U do not depend on θ. In our example, $C_L = -z_{\alpha/2}$ and $C_U = z_{\alpha/2}$. Finally, we must manipulate the inequalities in the probability statement so that

$$P[L(X_1, X_2, \ldots, X_n) \le \theta \le U(X_1, X_2, \ldots, X_n)] = 1 - \alpha \tag{8.10}$$

This gives $L(X_1, X_2, \ldots, X_n)$ and $U(X_1, X_2, \ldots, X_n)$ as the lower and upper confidence limits defining the $100(1 - \alpha)$ confidence interval for θ. The quantity $g(X_1, X_2, \ldots, X_n; \theta)$ is often called a *pivotal quantity* because we pivot on this quantity in Equation 8.9 to produce Equation 8.10. In our example, we manipulated the pivotal quantity $(\overline{X} - \mu)/(\sigma/\sqrt{n})$ to obtain $L(X_1, X_2, \ldots, X_n) = \overline{X} - z_{\alpha/2}\sigma/\sqrt{n}$ and $U(X_1, X_2, \ldots, X_n) = \overline{X} + z_{\alpha/2}\sigma/\sqrt{n}$.

8.1.5 Large-Sample Confidence Interval for μ

We have assumed that the population distribution is normal with unknown mean and known standard deviation σ. We now present a **large-sample CI** for μ that does not require these assumptions. Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with unknown mean μ and variance σ^2. Now if the sample size n is large, the central limit theorem implies that $\overline{X}$ has approximately a normal distribution with mean μ and variance σ^2/n. Therefore, $Z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ has approximately a standard normal distribution. This ratio could be used as a pivotal quantity and manipulated as in Section 8.1.1 to produce an approximate CI for μ. However, the standard deviation σ is unknown. It turns out that when n is large, replacing σ by the sample standard deviation S has little effect on the distribution of Z. This leads to the following useful result.

Large-Sample Confidence Interval on the Mean

When n is large, the quantity

$$\frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has an approximate standard normal distribution. Consequently,

$$\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}} \tag{8.11}$$

is a large-sample confidence interval for μ, with confidence level of approximately $100(1 - \alpha)\%$.

Equation 8.11 holds regardless of the shape of the population distribution. Generally, n should be at least 40 to use this result reliably. The central limit theorem generally holds for $n \ge 30$, but the larger sample size is recommended here because replacing s with S in Z results in additional variability.

EXAMPLE 8.4 | Mercury Contamination

An article in the 1993 volume of the *Transactions of the American Fisheries Society* reports the results of a study to investigate the mercury contamination in largemouth bass. A sample of fish was selected from 53 Florida lakes, and mercury concentration in the muscle tissue was measured (ppm). The mercury concentration values were

1.230	1.330	0.040	0.044	1.200	0.270
0.490	0.190	0.830	0.810	0.710	0.500
0.490	1.160	0.050	0.150	0.190	0.770
1.080	0.980	0.630	0.560	0.410	0.730
0.590	0.340	0.340	0.840	0.500	0.340
0.280	0.340	0.750	0.870	0.560	0.170
0.180	0.190	0.040	0.490	1.100	0.160
0.100	0.210	0.860	0.520	0.650	0.270
0.940	0.400	0.430	0.250	0.270	

The summary statistics for these data are as follows:

Variable	*N*	Mean	Median	StDev
Concentration	53	0.5250	0.4900	0.3486

Minimum	Maximum	*Q*1	*Q*3
0.0400	1.3300	0.2300	0.7900

Figure 8.3 presents the histogram and normal probability plot of the mercury concentration data. Both plots indicate that the distribution of mercury concentration is not normal and is positively skewed. We want to find an approximate 95% CI on μ. Because $n > 40$, the assumption of normality is not necessary to use in Equation 8.11. The required quantities are $n = 53, \bar{x} = 0.5250, s = 0.3486$, and $z_{0.025} = 1.96$. The approximate 95% CI on μ is

$$\bar{x} - z_{0.025}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + z_{0.025}\frac{s}{\sqrt{n}}$$

$$0.5250 - 1.96\frac{0.3486}{\sqrt{53}} \le \mu \le 0.5250 + 1.96\frac{0.3486}{\sqrt{53}}$$

$$0.4311 \le \mu \le 0.6189$$

Practical Interpretation: This interval is fairly wide because there is substantial variability in the mercury concentration measurements. A larger sample size would have produced a shorter interval.

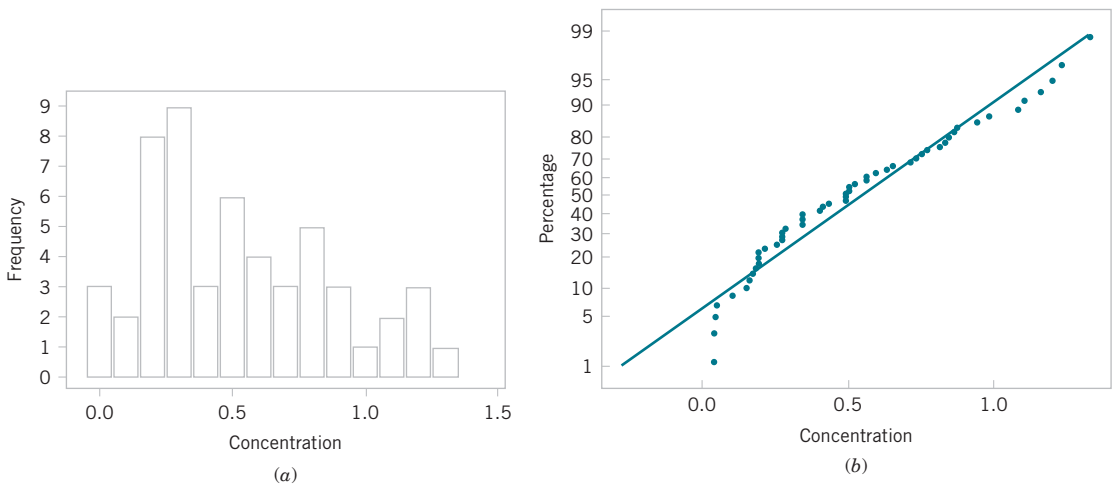


FIGURE 8.3

Mercury concentration in largemouth bass. (a) Histogram. (b) Normal probability plot.

Large-Sample Confidence Interval for a Parameter The large-sample confidence interval for μ in Equation 8.11 is a special case of a more general result. Suppose that θ is a parameter of a probability distribution, and let $\widehat{\Theta}$ be an estimator of θ. If $\widehat{\Theta}$ (1) has an approximate normal distribution, (2) is an approximately unbiased estimator for θ, and (3) has standard deviation $\sigma_{\widehat{\Theta}}$ that can be estimated from the sample data, the quantity $(\widehat{\Theta} - 0)/\sigma_{\widehat{\Theta}}$ has an approximate standard normal distribution. Then a large-sample approximate CI for θ is given by

Large-Sample Approximate Confidence Interval

$$\widehat{\theta} - z_{\alpha/2}\sigma_{\widehat{\Theta}} \le \theta \le \widehat{\theta} + z_{\alpha/2}\sigma_{\widehat{\Theta}} \tag{8.12}$$

Maximum likelihood estimators usually satisfy the three conditions just listed, so Equation 8.12 is often used when $\widehat{\Theta}$ is the maximum likelihood estimator of θ. Finally, note that Equation 8.12 can be used even when $\sigma_{\widehat{\Theta}}$ is a function of other unknown parameters (or of θ). Essentially, we simply use the sample data to compute estimates of the unknown parameters and substitute those estimates into the expression for $\sigma_{\widehat{\Theta}}$

8.2 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

When we are constructing confidence intervals on the mean μ of a normal population when σ^2 is known, we can use the procedure in Section 8.1.1. This CI is also approximately valid (because of the central limit theorem) regardless of whether or not the underlying population is normal so long as n is reasonably large ($n \ge 40$, say). As noted in Section 8.1.5, we can even handle the case of unknown variance for the large-sample-size situation. However, when the sample is small and σ^2 is unknown, we must make an assumption about the form of the underlying distribution to obtain a valid CI procedure. A reasonable assumption in many cases is that the underlying distribution is normal.

Many populations encountered in practice are well approximated by the normal distribution, so this assumption will lead to confidence interval procedures of wide applicability. In fact, moderate departure from normality will have little effect on validity. When the assumption is unreasonable, an alternative is to use nonparametric statistical procedures that are valid for any underlying distribution.

Suppose that the population of interest has a normal distribution with unknown mean μ and unknown variance σ^2. Assume that a random sample of size n, say, $X_1, X_2, \ldots, X_n$, is available, and let $\overline{X}$ and S^2 be the sample mean and variance, respectively.

We wish to construct a two-sided CI on μ. If the variance σ^2 is known, we know that $Z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ has a standard normal distribution. When σ^2 is unknown, a logical procedure is to replace σ with the sample standard deviation S. The random variable Z now becomes $T = (\overline{X} - \mu)/(S/\sqrt{n})$. A logical question is what effect replacing σ with S has on the distribution of the random variable T. If n is large, the answer to this question is "very little," and we can proceed to use the confidence interval based on the normal distribution from Section 8.1.5. However, n is usually small in most engineering problems, and in this situation, a different distribution must be employed to construct the CI.

8.2.1 *t* Distribution

_t_ Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. The random variable

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \tag{8.13}$$

has a t distribution with $n - 1$ degrees of freedom.

The t probability density function is

$$f(x) = \frac{\Gamma\left[(k+1)/2\right]}{\sqrt{\pi k}\,\Gamma(k/2)} \cdot \frac{1}{\left[(x^2/k) + 1\right]^{(k+1)/2}} \qquad -\infty < x < \infty \tag{8.14}$$

where k is the number of degrees of freedom. The mean and variance of the t distribution are zero and $k/(k-2)$ (for $k > 2$), respectively.

Several t distributions are shown in Figure 8.4. The general appearance of the t distribution is similar to the standard normal distribution in that both distributions are symmetric and unimodal, and the maximum ordinate value is reached when the mean $\mu = 0$. However, the t distribution has heavier tails than the normal; that is, it has more probability in the tails than does the normal distribution. As the number of degrees of freedom $k \to \infty$, the limiting form of the t distribution is the standard normal distribution. Generally, the number of degrees of freedom for t is the number of degrees of freedom associated with the estimated standard deviation.

Appendix Table V provides *percentage points* of the t distribution. We let $t_{\alpha,k}$ be the value of the random variable T with k degrees of freedom above which we find an area (or probability) α. Thus, $t_{\alpha,k}$ is an upper-tailed 100α percentage point of the t distribution with k degrees of freedom. This percentage point is shown in Figure 8.5. In Appendix Table V, the α values are the column headings, and the degrees of freedom are listed in the left column. To illustrate the use of the table, note that the t-value with 10 degrees of freedom having an area of 0.05 to the right is $t_{0.05,10} = 1.812$. That is,

$$P(T_{10} > t_{0.05,10}) = P(T_{10} > 1.812) = 0.05$$

Because the t distribution is symmetric about zero, we have $t_{1-\alpha,n} = -t_{\alpha,n}$; that is, the t-value having an area of $1 - \alpha$ to the right (and therefore an area of α to the left) is equal to the negative of the t-value that has area α in the right tail of the distribution. Therefore, $t_{0.95,10} = -t_{0.05,10} = -1.812$. Finally, because $t_{\alpha,\infty}$ is the standard normal distribution, the familiar z_α values appear in the last row of Appendix Table V.

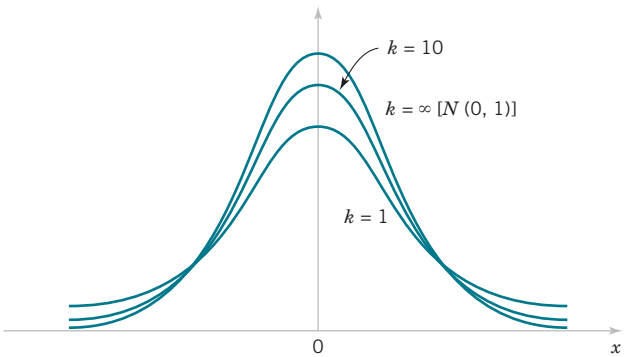


FIGURE 8.4
Probability density functions of several *t* distributions.

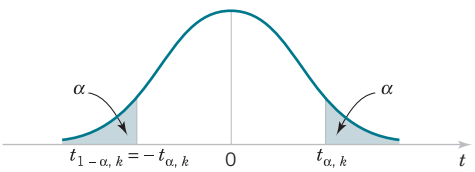


FIGURE 8.5
Percentage points of the *t* distribution.

8.2.2 *t* Confidence Interval on μ

It is easy to find a $100(1-\alpha)\%$ confidence interval on the mean of a normal distribution with unknown variance by proceeding essentially as we did in Section 8.1.1. We know that the distribution of $T = (\overline{X} - \mu)/(S/\sqrt{n})$ is t with $n-1$ degrees of freedom. Letting $t_{\alpha/2,n-1}$ be the upper $100\alpha/2$ percentage point of the t distribution with $n-1$ degrees of freedom, we may write

$$P(-t_{\alpha/2,n-1} \le T \le t_{\alpha/2,n-1}) = 1-\alpha$$

or

$$P\left(-t_{\alpha/2,n-1} \le \frac{\overline{X}-\mu}{S/\sqrt{n}} \le t_{\alpha/2,n-1}\right) = 1-\alpha$$

Rearranging this last equation yields

$$P(\overline{X} - t_{\alpha/2,n-1}S/\sqrt{n} \le \mu \le \overline{X} + t_{\alpha/2,n-1}S/\sqrt{n}) = 1-\alpha \tag{8.15}$$

This leads to the following definition of the $100(1-\alpha)\%$ two-sided confidence interval on μ.

Confidence Interval on the Mean, Variance Unknown

If $\overline{x}$ and s are the mean and standard deviation of a random sample from a normal distribution with unknown variance σ^2, a $100(1-\alpha)\%$ confidence interval on μ is given by

$$\overline{x} - t_{\alpha/2,n-1}s/\sqrt{n} \le \mu \le \overline{x} + t_{\alpha/2,n-1}s/\sqrt{n} \tag{8.16}$$

where $t_{\alpha/2,n-1}$ is the upper $100\alpha/2$ percentage point of the t distribution with $n-1$ degrees of freedom.

The assumption underlying this CI is that we are sampling from a normal population. However, the t distribution-based CI is relatively insensitive or robust to this assumption. Checking the normality assumption by constructing a normal probability plot of the data is a good general practice. Small to moderate departures from normality are not a cause for concern.

One-sided confidence bounds on the mean of a normal distribution are also of interest and are easy to find. Simply use only the appropriate lower or upper confidence limit from Equation 8.16 and replace $t_{\alpha/2,n-1}$ by $t_{\alpha,n-1}$.

EXAMPLE 8.5 | Alloy Adhesion

An article in the *Journal of Materials Engineering* ["Instrumented Tensile Adhesion Tests on Plasma Sprayed Thermal Barrier Coatings" (1989, Vol. 11(4), pp. 275–282)] describes the results of tensile adhesion tests on 22 U-700 alloy specimens. The load at specimen failure is as follows (in megapascals):

19.8	10.1	14.9	7.5	15.4	15.4
15.4	18.5	7.9	12.7	11.9	11.4
11.4	14.1	17.6	16.7	15.8	
19.5	8.8	13.6	11.9	11.4	

The sample mean is $\overline{x} = 13.71$, and the sample standard deviation is $s = 3.55$. Figures 8.6 and 8.7 show a box plot and a normal probability plot of the tensile adhesion test data, respectively. These displays provide good support for the assumption that the population is normally distributed. We want to find a 95% CI on μ. Since $n = 22$, we have $n-1 = 21$ degrees of freedom for t, so $t_{0.025,21} = 2.080$. The resulting CI is

$$\overline{x} - t_{\alpha/2,n-1}s/\sqrt{n} \le \mu \le \overline{x} + t_{\alpha/2,n-1}s/\sqrt{n}$$
$$13.71 - 2.080(3.55)/\sqrt{22} \le \mu \le 13.71 + 2.080(3.55)/\sqrt{22}$$
$$13.71 - 1.57 \le \mu \le 13.71 + 1.57$$
$$12.14 \le \mu \le 15.28$$

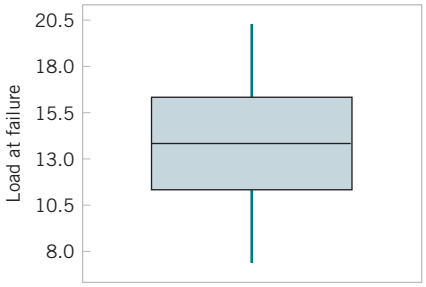


FIGURE 8.6

Box-and-whisker plot for the load at failure data.

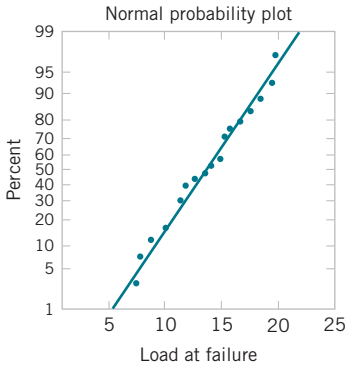


FIGURE 8.7

Normal probability plot of the load at failure data.

Practical Interpretation: The CI is fairly wide because there is a lot of variability in the tensile adhesion test measurements. A larger sample size would have led to a shorter interval.

It is not as easy to select a sample size n to obtain a specified length (or precision of estimation) for this CI as it was in the known-σ case because the length of the interval involves s (which is unknown before the data are collected), n, and $t_{\alpha/2,n-1}$. Note that the t-percentile depends on the sample size n. Consequently, an appropriate n can only be obtained through trial and error. The results of this will, of course, also depend on the reliability of our prior "guess" for σ.

8.3 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

Sometimes confidence intervals on the population variance or standard deviation are needed. When the population is modeled by a normal distribution, the tests and intervals described in this section are applicable. The following result provides the basis of constructing these confidence intervals.

χ^2 Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2, and let S^2 be the sample variance. Then the random variable

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \tag{8.17}$$

has a chi-square (χ^2) distribution with $n-1$ degrees of freedom.

The probability density function of a χ^2 random variable is

$$f(x) = \frac{1}{2^{k/2}\,\Gamma(k/2)}\, x^{(k/2)-1} e^{-x/2} \qquad x > 0 \tag{8.18}$$

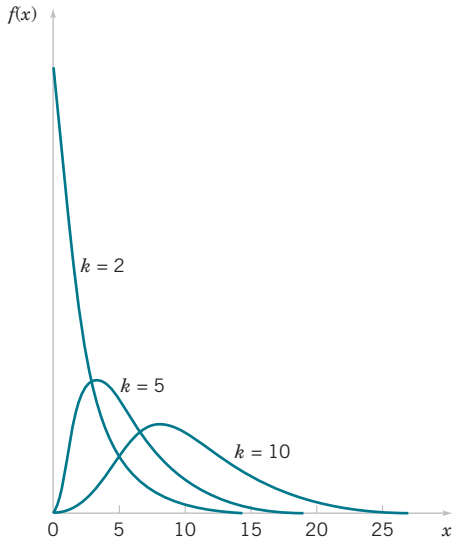


FIGURE 8.8

Probability density functions of several χ^2 distributions.

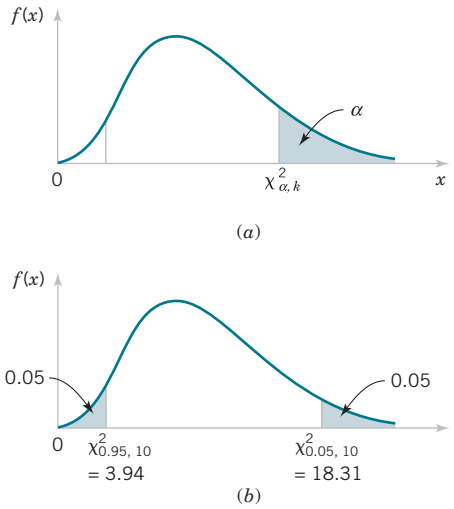


FIGURE 8.9

Percentage point of the χ^2 distribution. (a) The percentage point $\chi^2_{\alpha,k}$. (b) The upper percentage point $\chi^2_{0.05,10} = 18.31$ and the lower percentage point $\chi^2_{0.95,10} = 3.94$.

where k is the number of degrees of freedom. The mean and variance of the χ^2 distribution are k and $2k$, respectively. Several chi-square distributions are shown in Figure 8.8. Note that the chi-square random variable is non-negative and that the probability distribution is skewed to the right. However, as k increases, the distribution becomes more symmetric. As $k \to \infty$, the limiting form of the chi-square distribution is the normal distribution.

The *percentage points* of the χ^2 distribution are given in Table IV of the Appendix. Define $\chi^2_{\alpha,k}$ as the percentage point or value of the chi-square random variable with k degrees of freedom such that the probability that X^2 exceeds this value is α. That is,

$$P(X^2 > \chi^2_{\alpha,k}) = \int_{\chi^2_{\alpha,k}}^{\infty} f(u)\, du = \alpha$$

This probability is shown as the shaded area in Figure 8.9(a). To illustrate the use of Table IV, note that the areas α are the column headings and the degrees of freedom k are given in the left column. Therefore, the value with 10 degrees of freedom having an area (probability) of 0.05 to the right is $\chi^2_{0.05,10} = 18.31$. This value is often called an *upper* 5% point of chi-square with 10 degrees of freedom. We may write this as a probability statement as follows:

$$P(X^2 > \chi^2_{0.05,10}) = P(X^2 > 18.31) = 0.05$$

Conversely, a *lower* 5% point of chi-square with 10 degrees of freedom would be $\chi^2_{0.95,10} = 3.94$ (from Appendix A). Both of these percentage points are shown in Figure 8.9(b).

The construction of the $100(1 - \alpha)\%$ CI for σ^2 is straightforward. Because

$$X^2 = \frac{(n-1)S^2}{\sigma^2}$$

is chi-square with $n - 1$ degrees of freedom, we may write

$$P(\chi^2_{1-\alpha/2,n-1} \le X^2 \le \chi^2_{\alpha/2,n-1}) = 1 - \alpha$$

so that

$$P\left(\chi^2_{1-\alpha/2,n-1} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\alpha/2,n-1}\right) = 1-\alpha$$

This last equation can be rearranged as

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right) = 1-\alpha$$

This leads to the following definition of the confidence interval for σ^2.

Confidence Interval on the Variance

If s^2 is the sample variance from a random sample of n observations from a normal distribution with unknown variance σ^2, then a $100(1-\alpha)\%$ confidence interval on σ^2 is

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}} \tag{8.19}$$

where $\chi^2_{\alpha/2,n-1}$ and $\chi^2_{1-\alpha/2,n-1}$ are the upper and lower $100\alpha/2$ percentage points of the chi-square distribution with $n-1$ degrees of freedom, respectively. A confidence interval for σ has lower and upper limits that are the square roots of the corresponding limits in Equation 8.19.

It is also possible to find a $100(1-\alpha)\%$ lower confidence bound or upper confidence bound on σ^2.

One-Sided Confidence Bounds on the Variance

The $100(1-\alpha)\%$ lower and upper confidence bounds on σ^2 are

$$\frac{(n-1)s^2}{\chi^2_{\alpha,n-1}} \le \sigma^2 \quad \text{and} \quad \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha,n-1}} \tag{8.20}$$

respectively.

The CIs given in Equations 8.19 and 8.20 are less robust to the normality assumption. The distribution of $(n-1)S^2/\sigma^2$ can be very different from the chi-square if the underlying population is not normal.

EXAMPLE 8.6 | Detergent Filling

An automatic filling machine is used to fill bottles with liquid detergent. A random sample of 20 bottles results in a sample variance of fill volume of $s^2 = 0.0153^2$ (fluid ounce). If the variance of fill volume is too large, an unacceptable proportion of bottles will be under- or overfilled. We will assume that the fill volume is approximately normally distributed. A 95% upper confidence bound is found from Equation 8.20 as follows:

$$\sigma^2 \le \frac{(n-1)s^2}{\chi^2_{0.95,19}}$$

or

$$\sigma^2 \leq \frac{(19)0.0153}{10.117} = 0.0287 \text{ (fluid ounce)}^2$$

This last expression may be converted into a confidence interval on the standard deviation σ by taking the square root of both sides, resulting in

$$\sigma \leq 0.17$$

Practical Interpretation: Therefore, at the 95% level of confidence, the data indicate that the process standard deviation could be as large as 0.17 fluid ounce. The process engineer or manager now needs to determine whether a standard deviation this large could lead to an operational problem with under- or over-filled bottles.

8.4 Large-Sample Confidence Interval for a Population Proportion

It is often necessary to construct confidence intervals on a population proportion. For example, suppose that a random sample of size n has been taken from a large (possibly infinite) population and that $X(\leq n)$ observations in this sample belong to a class of interest. Then $\widehat{P} = X/n$ is a point estimator of the proportion of the population p that belongs to this class. Note that n and p are the parameters of a binomial distribution. Furthermore, from Chapter 4 we know that the sampling distribution of $\widehat{P}$ is approximately normal with mean p and variance $p(1-p)/n$, if p is not too close to either 0 or 1 and if n is relatively large. Typically, to apply this approximation we require that np and $n(1-p)$ be greater than or equal to 5. We use the normal approximation in this section.

Normal Approximation for a Binomial Proportion

If n is large, the distribution of

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\widehat{P} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

is approximately standard normal.

To construct the confidence interval on p, note that

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) \simeq 1 - \alpha$$

so

$$P\left(-z_{\alpha/2} \leq \frac{\widehat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{\alpha/2}\right) \simeq 1 - \alpha$$

This may be rearranged as

$$P\left(\widehat{P} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \widehat{P} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) \simeq 1 - \alpha \tag{8.21}$$

The quantity $\sqrt{p(1-p)/n}$ in Equation 8.21 is called the *standard error of the point estimator* $\widehat{P}$. This was discussed in Chapter 7. Unfortunately, the upper and lower limits of the confidence interval obtained from Equation 8.21 contain the unknown parameter p. However, as suggested

at the end of Section 8.1.5, a solution that is often satisfactory is to replace p by $\widehat{P}$ in the standard error, which results in

$$P\left(\widehat{P} - z_{\alpha/2}\sqrt{\frac{\widehat{P}(1-\widehat{P})}{n}} \le p \le \widehat{P} + z_{\alpha/2}\sqrt{\frac{\widehat{P}(1-\widehat{P})}{n}}\right) \simeq 1 - \alpha \tag{8.22}$$

This leads to the approximate $100(1-\alpha)\%$ confidence interval on p.

Approximate Confidence Interval on a Binomial Proportion

If $\widehat{p}$ is the proportion of observations in a random sample of size n that belongs to a class of interest, an approximate $100(1-\alpha)\%$ confidence interval on the proportion p of the population that belongs to this class is

$$\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \le p \le \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \tag{8.23}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution.

This procedure depends on the adequacy of the normal approximation to the binomial. To be reasonably conservative, this requires that np and $n(1-p)$ be greater than or equal to 5. In situations when this approximation is inappropriate, particularly in cases when n is small, other methods must be used. Tables of the binomial distribution could be used to obtain a confidence interval for p. However, we could also use numerical methods that are implemented on the binomial probability mass function in some computer programs.

EXAMPLE 8.7 | Crankshaft Bearings

In a random sample of 85 automobile engine crankshaft bearings, 10 have a surface finish that is rougher than the specifications allow. Therefore, a point estimate of the proportion of bearings in the population that exceeds the roughness specification is $\widehat{p} = x/n = 10/85 = 0.12$. A 95% two-sided confidence interval for p is computed from Equation 8.23 as

$$\widehat{p} - z_{0.025}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \le p \le \widehat{p} + z_{0.025}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

or

$$0.12 - 1.96\sqrt{\frac{0.12(0.88)}{85}} \le p \le 0.12 + 1.96\sqrt{\frac{0.12(0.88)}{85}}$$

which simplifies to

$$0.0509 \le p \le 0.2243$$

Practical Interpretation: This is a wide CI. Although the sample size does not appear to be small ($n = 85$), the value of $\widehat{p}$ is fairly small, which leads to a large standard error for $\widehat{p}$ contributing to the wide CI.

Choice of Sample Size Because $\widehat{P}$ is the point estimator of p, we can define the error in estimating p by $\widehat{P}$ as $E = |p - \widehat{P}|$. Note that we are approximately $100(1-\alpha)\%$ confident that this error is less than $z_{\alpha/2}\sqrt{p(1-p)/n}$. For instance, in Example 8.7, we are 95% confident that the sample proportion $\widehat{p} = 0.12$ differs from the true proportion p by an amount not exceeding 0.07.

In situations when the sample size can be selected, we may choose n to be $100(1-\alpha)\%$ confident that the error is less than some specified value E. If we set $E = z_{\alpha/2}\sqrt{p(1-p)/n}$ and solve for n, the appropriate sample size is

Sample Size for a Specified Error on a Binomial Proportion

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p) \tag{8.24}$$

An estimate of p is required to use Equation 8.24. If an estimate $\hat{p}$ from a previous sample is available, it can be substituted for p in Equation 8.24, or perhaps a subjective estimate can be made. If these alternatives are unsatisfactory, a preliminary sample can be taken, $\hat{p}$ computed, and then Equation 8.24 used to determine how many additional observations are required to estimate p with the desired accuracy. Another approach to choosing n uses the fact that the sample size from Equation 8.24 will always be a maximum for $p = 0.5$ [that is, $p(1-p) \le 0.25$ with equality for $p = 0.5$], and this can be used to obtain an upper bound on n. In other words, we are at least $100(1-\alpha)\%$ confident that the error in estimating p by $\hat{P}$ is less than E if the sample size is

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 (0.25) \tag{8.25}$$

EXAMPLE 8.8 | Crankshaft Bearings

Consider the situation in Example 8.7. How large a sample is required if we want to be 95% confident that the error in using $\hat{p}$ to estimate p is less than 0.05? Using $\hat{p} = 0.12$ as an initial estimate of p, we find from Equation 8.24 that the required sample size is

$$n = \left(\frac{z_{0.025}}{E}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{1.96}{0.05}\right)^2 0.12(0.88) \cong 163$$

If we wanted to be *at least* 95% confident that our estimate $\hat{p}$ of the true proportion p was within 0.05 regardless of the value of p, we would use Equation 8.25 to find the sample size

$$n = \left(\frac{z_{0.025}}{E}\right)^2 (0.25) = \left(\frac{1.96}{0.05}\right)^2 (0.25) \cong 385$$

Practical Interpretation: Notice that if we have information concerning the value of p, either from a preliminary sample or from past experience, we could use a smaller sample while maintaining both the desired precision of estimation and the level of confidence.

One-Sided Confidence Bounds We may find approximate one-sided confidence bounds on p by using a simple modification of Equation 8.23.

Approximate One-Sided Confidence Bounds on a Binomial Proportion

The approximate $100(1-\alpha)\%$ lower and upper confidence bounds are

$$\hat{p} - z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \quad \text{and} \quad p \le \hat{p} + z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{8.26}$$

respectively.

A Different Confidence Interval on the Binomial Proportion There is a different way to construct a CI on a binomial proportion than the traditional approach in Equation 8.23. Starting with Equation 8.22 and replacing the inequalities with an equality and solving the resulting quadratic equation for p results in

$$p = \frac{\hat{p} + \frac{z^2_{\alpha/2}}{2n} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2_{\alpha/2}}{4n^2}}}{1 + z^2_{\alpha/2}/n}$$

This implies that a two-sided CI on a proportion p is as follows:

$$UCL = \frac{\hat{p} + \frac{z^2_{\alpha/2}}{2n} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2_{\alpha/2}}{4n^2}}}{1 + z^2_{\alpha/2}/n}$$

$$LCL = \frac{\hat{p} + \frac{z^2_{\alpha/2}}{2n} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2_{\alpha/2}}{4n^2}}}{1 + z^2_{\alpha/2}/n} \tag{8.27}$$

The article by Agresti and Coull in *The American Statistician* ("Approximate Better Than 'Exact' for Interval Estimation of a Binomial Proportion," 1998, Vol. 52, pp. 119–126) reports that the actual confidence level for the CI in Equation 8.27 is closer to the "advertised" or nominal level for almost all values of α and p than for the traditional CI in Equation 8.23. The authors also report that this new interval can be used with nearly all sample sizes. So the requirements that $n\hat{p} \geq 5$ or 10 or $n(1-\hat{p}) \geq 5$ or 10 are not too important. If the sample size is large, the quantity $z^2_{\alpha/2}/(2n)$ will be small relative to $\hat{p}$, $z^2_{\alpha/2}/(4n^2)$ will be small relative to $\hat{p}(1-\hat{p})/n$, and $z^2_{\alpha/2}/n$ will be small, so as a result the **Agresti-Coull CI** in Equation 8.27 will reduce to the traditional CI given in Equation 8.23.

8.5 Guidelines for Constructing Confidence Intervals

The most difficult step in constructing a confidence interval is often the match of the appropriate calculation to the objective of the study. Common cases are listed in Table 8.1 along with the reference to the section that covers the appropriate calculation for a confidence interval test. Table 8.1 provides a simple road map to help select the appropriate analysis. Two primary comments can help identify the analysis:

1. Determine the parameter (and the distribution of the data) that will be bounded by the confidence interval or tested by the hypothesis.
2. Check if other parameters are known or need to be estimated.

In Chapter 9, we study a procedure closely related to confidence intervals called *hypothesis testing*. Table 8.1 can be used for those procedures also. This road map is extended to more cases in Chapter 10.

TABLE 8.1 The Roadmap for Constructing Confidence Intervals and Performing Hypothesis Tests, One-Sample Case

Parameter to Be Bounded by the Confidence Interval or Tested with a Hypothesis?	Symbol	Other Parameters?	Confidence Interval Section	Hypothesis Test Section	Comments
Mean of normal distribution	μ	Standard deviation σ known	8.1	9.2	Large sample size is often taken to be $n \geq 40$
Mean of arbitrary distribution with large sample size	μ	Sample size large enough that central limit theorem applies and σ is essentially known	8.1.5	9.2.3	
Mean of normal distribution	μ	Standard deviation σ unknown and estimated	8.2	9.3	
Variance (or standard deviation) of normal distribution	σ^2	Mean μ unknown and estimated	8.3	9.4	
Population proportion	p	None	8.4	9.5	

8.6 Bootstrap Confidence Interval

In Section 7.3.4, we saw how a computer-intensive technique called the *bootstrap* could be used to find the estimated standard error of a statistic, say $\hat{\theta}$. The bootstrap technique can also be used to find confidence intervals. These techniques can be useful in situations in which a "standard" CI is not readily available. To illustrate the general approach, let's consider a case for which there is a standard CI, the $100(1-\alpha)\%$ CI on the mean of a normal distribution with known variance. Here the parameter of interest is the population mean μ, and the statistic that estimates μ is the sample average $\overline{X}$. The quantity $z_{\alpha/2}\sigma/\sqrt{n}$ is the $100(1-\alpha/2)$ percentile of the distribution of $\hat{\theta}_i^B - \overline{\theta}^B, i = 1, 2, \ldots, n_B$ and by the same logic, the quantity $-z_{\alpha/2}\sigma/\sqrt{n}$ is the $100(\alpha/2)$ percentile of the distribution of $\overline{X} - \mu$. Therefore, the $100(1-\alpha/2)\%$ CI can be written as:

$$P(\alpha/2\text{th percentile} \leq \overline{X} - \mu \leq (1-\alpha/2)\text{th percentile}) = 1 - \alpha/2$$

This can be rearranged as

$$P(\overline{X} - (1-\alpha/2\text{th}) \text{ percentile} \leq \mu \leq \overline{X} + \alpha/2\text{th percentile}) = 1 - \alpha/2$$

So the lower confidence bound is $\overline{X} - (1-\alpha/2)$th percentile of the distribution of $\overline{X} - \mu$ and the upper confidence bound is $\overline{X} + \alpha/2$th percentile of the distribution of $\overline{X} - \mu$. When these percentiles cannot be easily determined for some arbitrary parameter θ, they can often be estimated by using bootstrap samples. The procedure would consist of taking n_B bootstrap samples, calculating the bootstrap estimates $\hat{\theta}_1^B, \hat{\theta}_2^B, \ldots, \hat{\theta}_{n_B}^B$ and $\overline{\theta}^B$, and then computing the differences $\hat{\theta}_i^B - \overline{\theta}^B, i = 1, 2, \ldots, n_B$. The $\alpha/2$ smallest and largest of these differences are the estimates of the percentiles required to construct the bootstrap CI.

8.7 Tolerance and Prediction Intervals

8.7.1 Prediction Interval for a Future Observation

In some problem situations, we may be interested in predicting a future observation of a variable. This is a different problem than estimating the mean of that variable, so a confidence interval is

not appropriate. In this section, we show how to obtain a $100(1-\alpha)\%$ **prediction interval** on a future value of a normal random variable.

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a normal population. We wish to predict the value X_{n+1}, a single **future** observation. A point prediction of X_{n+1} is $\overline{X}$, the sample mean. The prediction error is $X_{n+1} - \overline{X}$. The expected value of the prediction error is

$$E(X_{n+1} - \overline{X}) = \mu - \mu = 0$$

and the variance of the prediction error is

$$V(X_{n+1} - \overline{X}) = \sigma^2 + \frac{\sigma^2}{n} = \sigma^2\left(1 + \frac{1}{n}\right)$$

because the future observation X_{n+1} is independent of the mean of the current sample $\overline{X}$. The prediction error $X_{n+1} - \overline{X}$ is normally distributed. Therefore,

$$Z = \frac{X_{n+1} - \overline{X}}{\sigma\sqrt{1 + \frac{1}{n}}}$$

has a standard normal distribution. Replacing σ with S results in

$$T = \frac{X_{n+1} - \overline{X}}{S\sqrt{1 + \frac{1}{n}}}$$

which has a t distribution with $n - 1$ degrees of freedom. Manipulating T as we have done previously in the development of a CI leads to a prediction interval on the future observation X_{n+1}.

Prediction Interval

A $100(1-\alpha)\%$ **prediction interval (PI)** on a single future observation from a normal distribution is given by

$$\overline{x} - t_{\alpha/2,n-1}s\sqrt{1 + \frac{1}{n}} \le X_{n+1} \le \overline{x} + t_{\alpha/2,n-1}\,s\sqrt{1 + \frac{1}{n}} \tag{8.28}$$

The prediction interval for X_{n+1} will always be longer than the confidence interval for μ because more variability is associated with the prediction error than with the error of estimation. This is easy to see because the prediction error is the difference between two random variables $(X_{n+1} - \overline{X})$, and the estimation error in the CI is the difference between one random variable and a constant $(\overline{X} - \mu)$. As n gets larger $(n \to \infty)$, the length of the CI decreases to zero, essentially becoming the single value μ, but the length of the PI approaches $2z_{\alpha/2}\sigma$. So as n increases, the uncertainty in estimating μ goes to zero, although there will always be uncertainty about the future value X_{n+1}, even when estimating any of the distribution parameters is not necessary.

We noted in Section 8.2 that the t distribution-based CI for μ was robust to the normality assumption when n is small. The practical implication of this is that although we have computed a 95% CI, the actual confidence level will not be exactly 95%, but it will be very close—maybe 93% or 94%. Prediction intervals, on the other hand, are very sensitive to the normality assumption, and Equation 8.28 should not be used unless we are very comfortable with the normality assumption.

EXAMPLE 8.9 | Alloy Adhesion

Reconsider the tensile adhesion tests on specimens of U-700 alloy described in Example 8.5. The load at failure for $n = 22$ specimens was observed, and we found that $\bar{x} = 13.71$ and $s = 3.55$. The 95% confidence interval on μ was $12.14 \le \mu \le 15.28$. We plan to test a 23rd specimen. A 95% prediction interval on the load at failure for this specimen is

$$\bar{x} - t_{\alpha/2,n-1}\, s\sqrt{1+\frac{1}{n}} \le X_{n+1} \le \bar{x} + t_{\alpha/2,n-1}\, s\sqrt{1+\frac{1}{n}}$$

$$13.71 - (2.080)3.55\sqrt{1+\frac{1}{22}} \le X_{23} \le 13.71 + (2.080)3.55\sqrt{1+\frac{1}{22}}$$

$$6.16 \le X_{23} \le 21.26$$

Practical Interpretation: Notice that the prediction interval is considerably longer than the CI. This is because the CI is an estimate of a parameter, but the PI is an interval estimate of a single future observation.

8.7.2 Tolerance Interval for a Normal Distribution

Consider a population of semiconductor processors. Suppose that the speed of these processors has a normal distribution with mean $\mu = 600$ megahertz and standard deviation $\sigma = 30$ megahertz. Then the interval from $600 - 1.96(30) = 541.2$ to $600 + 1.96(30) = 658.8$ megahertz captures the speed of 95% of the processors in this population because the interval from -1.96 to 1.96 captures 95% of the area under the standard normal curve. The interval from $\mu - z_{\alpha/2}\sigma$ to $\mu + z_{\alpha/2}\sigma$ is called a **tolerance interval**.

If μ and σ are unknown, we can use the data from a random sample of size n to compute $\bar{x}$ and s and then form the interval $(\bar{x} - 1.96s, \bar{x} + 1.96s)$. However, because of sampling variability in $\bar{x}$ and s, it is likely that this interval will contain less than 95% of the values in the population. The solution to this problem is to replace 1.96 with some value that will make the proportion of the distribution contained in the interval 95% with some level of confidence. Fortunately, it is easy to do this.

Tolerance Interval

A **tolerance interval** for capturing at least γ% of the values in a normal distribution with confidence level $100(1-\alpha)$% is

$$\bar{x} - ks, \qquad \bar{x} + ks$$

where k is a tolerance interval factor found in Appendix Table XII. Values are given for $\gamma = 90\%, 95\%$, and 99%, and for 90%, 95%, and 99% confidence.

This interval is very sensitive to the normality assumption. One-sided tolerance bounds can also be computed. The tolerance factors for these bounds are also given in Appendix Table XII.

EXAMPLE 8.10 | Alloy Adhesion

Let's reconsider the tensile adhesion tests originally described in Example 8.5. The load at failure for $n = 22$ specimens was observed, and we found that $\bar{x} = 13.71$ and $s = 3.55$. We want to find a tolerance interval for the load at failure that includes 90% of the values in the population with 95% confidence. From Appendix Table XII, the tolerance factor k for $n = 22$, $\gamma = 0.90$, and 95% confidence is $k = 2.264$. The desired tolerance interval is

$$(\bar{x} - ks, \bar{x} + ks)$$

or

$$[13.71 - (2.264)3.55, 13.71 + (2.264)3.55]$$

which reduces to (5.67, 21.74).

Practical Interpretation: We can be 95% confident that at least 90% of the values of load at failure for this particular alloy lie between 5.67 and 21.74 megapascals.

From Appendix Table XII, we note that as $n \to \infty$, the value of k goes to the z-value associated with the desired level of containment for the normal distribution. For example, if we want 90% of the population to fall in the two-sided tolerance interval, k approaches $z_{0.05} = 1.645$ as $n \to \infty$. Note that as $n \to \infty$, a $100(1 - \alpha)\%$ prediction interval on a future value approaches a tolerance interval that contains $100(1 - \alpha)\%$ of the distribution.

Important Terms and Concepts

Agresti-Coull confidence interval on a population proportion
Chi-square distribution
Confidence coefficient
Confidence interval
Confidence interval on a population proportion
Confidence interval on the variance of a normal distribution
Confidence interval on the mean of a normal distribution
Confidence level
Error in estimation
Large-sample confidence interval
One-sided confidence bound
Prediction interval
t distribution
Tolerance interval
Two-sided confidence interval

CHAPTER 9

Tests of Hypotheses for a Single Sample

Alain Nogues/Sygma/Sygma/Getty Images

LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Structure engineering decision-making problems as hypothesis tests
2. Test hypotheses on the mean of a normal distribution using either a *Z*-test or a *t*-test procedure
3. Test hypotheses on the variance or standard deviation of a normal distribution
4. Test hypotheses on a population proportion
5. Use the *P*-value approach for making decisions in hypothesis tests
6. Compute power and type II error probability, and make sample size selection decisions for tests on means, variances, and proportions
7. Explain and use the relationship between confidence intervals and hypothesis tests
8. Use the chi-square goodness-of-fit test to check distributional assumptions
9. Apply contingency table tests
10. Apply nonparametric tests
11. Use equivalence testing
12. Combine *P*-values

CHAPTER OUTLINE

9.1 Hypothesis Testing
- 9.1.1 Statistical Hypotheses
- 9.1.2 Tests of Statistical Hypotheses
- 9.1.3 One-Sided and Two-Sided Hypotheses
- 9.1.4 *P*-Values in Hypothesis Tests
- 9.1.5 Connection between Hypothesis Tests and Confidence Intervals
- 9.1.6 General Procedure for Hypothesis Tests

9.2 Tests on the Mean of a Normal Distribution, Variance Known
- 9.2.1 Hypothesis Tests on the Mean
- 9.2.2 Type II Error and Choice of Sample Size
- 9.2.3 Large-Sample Test

9.3 Tests on the Mean of a Normal Distribution, Variance Unknown
- 9.3.1 Hypothesis Tests on the Mean
- 9.3.2 Type II Error and Choice of Sample Size

9.4 Tests on the Variance and Standard Deviation of a Normal Distribution
- 9.4.1 Hypothesis Tests on the Variance
- 9.4.2 Type II Error and Choice of Sample Size

9.5 Tests on a Population Proportion
- 9.5.1 Large-Sample Tests on a Proportion
- 9.5.2 Type II Error and Choice of Sample Size

9.6 Summary Table of Inference Procedures for a Single Sample

9.7 Testing for Goodness of Fit

9.8 Contingency Table Tests

9.9 Nonparametric Procedures
- 9.9.1 The Sign Test
- 9.9.2 The Wilcoxon Signed-Rank Test
- 9.9.3 Comparison to the *t*-Test

9.10 Equivalence Testing

9.11 Combining *P*-Values

Introduction

In the previous two chapters, we showed how a parameter of a population can be estimated from sample data, using either a *point estimate* (Chapter 7) or an interval of likely values called a *confidence interval* (Chapter 8). In many situations, a different type of problem is of interest; there are two competing claims about the value of a parameter, and the engineer must determine which claim is correct. For example, suppose that an engineer is designing an air crew escape system that consists of an ejection seat and a rocket motor that powers the seat. The rocket motor contains a propellant, and for the ejection seat to function properly, the propellant should have a mean burning rate of 50 cm/sec. If the burning rate is too low, the ejection seat may not function properly, leading to an unsafe ejection and possible injury of the pilot. Higher burning rates may imply instability in the propellant or an ejection seat that is too powerful, again leading to possible pilot injury. So the practical engineering question that must be answered is: Does the mean burning rate of the propellant equal 50 cm/sec, or is it some other value (either higher or lower)? This type of question can be answered using a statistical technique called **hypothesis testing**. This chapter focuses on the basic principles of hypothesis testing and provides techniques for solving the most common types of hypothesis testing problems involving a single sample of data.

9.1 Hypothesis Testing

9.1.1 Statistical Hypotheses

In the previous chapter, we illustrated how to construct a confidence interval estimate of a parameter from sample data. However, many problems in engineering require that we decide which of two competing claims or statements about some parameter is true. The statements are called

hypotheses, and the decision-making procedure is called **hypothesis testing**. This is one of the most useful aspects of statistical inference, because many types of decision-making problems, tests, or experiments in the engineering world can be formulated as hypothesis-testing problems. Furthermore, as we will see, a very close connection exists between hypothesis testing and confidence intervals.

Statistical hypothesis testing and confidence interval estimation of parameters are the fundamental methods used at the data analysis stage of a *comparative experiment* in which the engineer is interested, for example, in comparing the mean of a population to a specified value. These simple comparative experiments are frequently encountered in practice and provide a good foundation for the more complex experimental design problems that we discuss in Chapters 13 and 14. In this chapter, we discuss comparative experiments involving a single population, and our focus is on testing hypotheses concerning the parameters of the population.

We now give a formal definition of a statistical hypothesis.

Statistical Hypothesis

A **statistical hypothesis** is a statement about the parameters of one or more populations.

Because we use probability distributions to represent populations, a statistical hypothesis may also be thought of as a statement about the probability distribution of a random variable. The hypothesis will usually involve one or more parameters of this distribution.

For example, consider the air crew escape system described in the introduction. Suppose that we are interested in the burning rate of the solid propellant. Burning rate is a random variable that can be described by a probability distribution. Suppose that our interest focuses on the mean burning rate (a parameter of this distribution). Specifically, we are interested in deciding whether or not the mean burning rate is 50 centimeters per second. We may express this formally as

$$H_0\colon \mu = 50 \text{ centimeters per second} \qquad H_1\colon \mu \neq 50 \text{ centimeters per second} \tag{9.1}$$

The statement $H_0\colon \mu = 50$ centimeters per second in Equation 9.1 is called the **null hypothesis**. This is a claim that is initially assumed to be true. The statement $H_1\colon \mu \neq 50$ centimeters per second is called the **alternative hypothesis** and it is a statement that condradicts the null hypothesis. Because the alternative hypothesis specifies values of μ that could be either greater or less than 50 centimeters per second, it is called a **two-sided alternative hypothesis**. In some situations, we may wish to formulate a **one-sided alternative hypothesis**, as in

$$\begin{gathered} H_0\colon \mu = 50 \text{ centimeters per second} \qquad H_1\colon \mu < 50 \text{ centimeters per second} \\ \text{or} \\ H_0\colon \mu = 50 \text{ centimeters per second} \qquad H_1\colon \mu > 50 \text{ centimeters per second} \end{gathered} \tag{9.2}$$

We will always state the null hypothesis as an equality claim. However, when the alternative hypothesis is stated with the $<$ sign, the implicit claim in the null hypothesis can be taken as $\geq$ and when the alternative hypothesis is stated with the $>$ sign, the implicit claim in the null hypothesis can be taken as $\leq$.

It is important to remember that hypotheses are always statements about the population or distribution under study, not statements about the sample. The value of the population parameter specified in the null hypothesis (50 centimeters per second in the preceding example) is usually determined in one of three ways. First, it may result from past experience or knowledge of the process or even from previous tests or experiments. The objective of hypothesis testing, then, is usually to determine whether the parameter value has changed. Second, this value may be determined from some theory or model regarding the process under study. Here the objective

of hypothesis testing is to verify the theory or model. A third situation arises when the value of the population parameter results from external considerations, such as design or engineering specifications, or from contractual obligations. In this situation, the usual objective of hypothesis testing is conformance testing.

A procedure leading to a decision about the null hypothesis is called a **test of a hypothesis**. Hypothesis-testing procedures rely on using the information in a random sample from the population of interest. If this information is consistent with the null hypothesis, we will not reject it; however, if this information is inconsistent with the null hypothesis, we will conclude that the null hypothesis is false and reject it in favor of the alternative. We emphasize that the truth or falsity of a particular hypothesis can never be known with certainty unless we can examine the entire population. This is usually impossible in most practical situations. Therefore, a hypothesis-testing procedure should be developed with the probability of reaching a wrong conclusion in mind. Testing the hypothesis involves taking a random sample, computing a **test statistic** from the sample data, and then using the test statistic to make a decision about the null hypothesis.

9.1.2 Tests of Statistical Hypotheses

To illustrate the general concepts, consider the propellant burning rate problem introduced earlier. The null hypothesis is that the mean burning rate is 50 centimeters per second, and the alternate is that it is not equal to 50 centimeters per second. That is, we wish to test

$$H_0: \mu = 50 \text{ centimeters per second}$$
$$H_1: \mu \neq 50 \text{ centimeters per second}$$

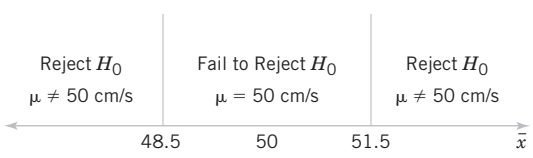
Suppose that a sample of $n = 10$ specimens is tested and that the sample mean burning rate $\bar{x}$ is observed. The sample mean is an estimate of the true population mean μ. A value of the sample mean $\bar{x}$ that falls close to the hypothesized value of $\mu = 50$ centimeters per second does not conflict with the null hypothesis that the true mean μ is really 50 centimeters per second. On the other hand, a sample mean that is considerably different from 50 centimeters per second is evidence in support of the alternative hypothesis H_1. Thus, the sample mean is the test statistic in this case.

The sample mean can take on many different values. Suppose that if $48.5 \leq \bar{x} \leq 51.5$, we will not reject the null hypothesis H_0: $\mu = 50$, and if either $\bar{x} < 48.5$ or $\bar{x} > 51.5$, we will reject the null hypothesis in favor of the alternative hypothesis H_1: $\mu \neq 50$. This is illustrated in Figure 9.1. The values of $\bar{x}$ that are less than 48.5 and greater than 51.5 constitute the **critical region** for the test; all values that are in the interval $48.5 \leq \bar{x} \leq 51.5$ form a region for which we will fail to reject the null hypothesis. By convention, this is usually called the **acceptance region**. The boundaries between the critical regions and the acceptance region are called the **critical values**. In our example, the critical values are 48.5 and 51.5. It is customary to state conclusions relative to the null hypothesis H_0. Therefore, we reject H_0 in favor of H_1 if the test statistic falls in the critical region and fails to reject H_0 otherwise.

This decision procedure can lead to either of two wrong conclusions. For example, the true mean burning rate of the propellant could be equal to 50 centimeters per second. However, for the randomly selected propellant specimens that are tested, we could observe a value of the test statistic $\bar{x}$ that falls into the critical region. We would then reject the null hypothesis H_0 in favor of the alternate H_1 when, in fact, H_0 is really true. This type of wrong conclusion is called a **type I error**.

FIGURE 9.1

Decision criteria for testing H_0: $\mu = 50$ centimeters per second versus H_1: $\mu \neq 50$ centimeters per second.



Type I Error

Rejecting the null hypothesis H_0 when it is true is defined as a **type I error**.

Now suppose that the true mean burning rate is different from 50 centimeters per second, yet the sample mean $\bar{x}$ falls in the acceptance region. In this case, we would fail to reject H_0 when it is false, and this leads to the other type of error.

Type II Error

Failing to reject the null hypothesis when it is false is defined as a **type II error**.

Thus, in testing any statistical hypothesis, four different situations determine whether the final decision is correct or in error. These situations are presented in Table 9.1.

Because our decision is based on random variables, probabilities can be associated with the type I and type II errors in Table 9.1. The probability of making a type I error is denoted by the Greek letter α.

Probability of Type I Error

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true}) \tag{9.3}$$

Sometimes the type I error probability is called the **significance level**, the **α-error**, or the **size of the test**. In the propellant burning rate example, a type I error will occur when either $\bar{x} > 51.5$ or $\bar{x} < 48.5$ when the true mean burning rate really is $\mu = 50$ centimeters per second. Suppose that the standard deviation of the burning rate is $\sigma = 2.5$ centimeters per second and that the burning rate has a distribution for which the conditions of the central limit theorem apply, so the distribution of the sample mean is approximately normal with mean $\mu = 50$ and standard deviation $\sigma/\sqrt{n} = 2.5/\sqrt{10} = 0.79$. The probability of making a type I error (or the significance level of our test) is equal to the sum of the areas that have been shaded in the tails of the normal distribution in Figure 9.2. We may find this probability as

$$\alpha = P(\bar{X} < 48.5 \text{ when } \mu = 50) + P(\bar{X} > 51.5 \text{ when } \mu = 50)$$

The z-values that correspond to the critical values 48.5 and 51.5 are

$$z_1 = \frac{48.5 - 50}{0.79} = -1.90 \quad \text{and} \quad z_2 = \frac{51.5 - 50}{0.79} = 1.90$$

Therefore,

$$\alpha = P(z < -1.90) + P(z > 1.90) = 0.0287 + 0.0287 = 0.0574$$

TABLE 9.1 Decisions in Hypothesis Testing

Decision	H_0 Is True	H_0 Is False
Fail to reject H_0	No error	Type II error
Reject H_0	Type I error	No error

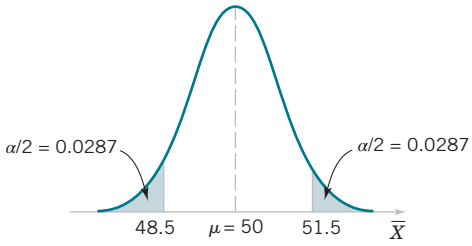


FIGURE 9.2
The critical region for H_0: $\mu = 50$ versus H_1: $\mu \neq 50$ and $n = 10$.

This is the type I error probability. This implies that 5.74% of all random samples would lead to rejection of the hypothesis H_0: $\mu = 50$ centimeters per second when the true mean burning rate is really 50 centimeters per second.

From an inspection of Figure 9.2, notice that we can reduce α by widening the acceptance region. For example, if we make the critical values 48 and 52, the value of α is

$$\alpha = P\left(z < -\frac{48-50}{0.79}\right) + P\left(z > \frac{52-50}{0.79}\right) = P(z < -2.53) + P(z > 2.53)$$
$$= 0.0057 + 0.0057 = 0.0114$$

We could also reduce α by increasing the sample size. If $n = 16$, $\sigma/\sqrt{n} = 2.5/\sqrt{16} = 0.625$ and using the original critical region from Figure 9.1, we find

$$z_1 = \frac{48.5-50}{0.625} = -2.40 \qquad \text{and} \qquad z_2 = \frac{51.5-50}{0.625} = 2.40$$

Therefore,

$$\alpha = P(Z < -2.40) + P(Z > 2.40) = 0.0082 + 0.0082 = 0.0164$$

In evaluating a hypothesis-testing procedure, it is also important to examine the probability of a type II error, which we denote by β. That is,

Probability of Type II Error

$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false}) \qquad (9.4)$$

To calculate β (sometimes called the **β-error**), we must have a specific alternative hypothesis; that is, we must have a particular value of μ. For example, suppose that it is important to reject the null hypothesis H_0: $\mu = 50$ whenever the mean burning rate μ is greater than 52 centimeters per second or less than 48 centimeters per second. We could calculate the probability of a type II error β for the values $\mu = 52$ and $\mu = 48$ and use this result to tell us something about how the test procedure would perform. Specifically, how will the test procedure work if we wish to detect, that is, reject H_0, for a mean value of $\mu = 52$ or $\mu = 48$? Because of symmetry, it is necessary to evaluate only one of the two cases—say, find the probability of failing to reject the null hypothesis H_0: $\mu = 50$ centimeters per second when the true mean is $\mu = 52$ centimeters per second.

Figure 9.3 helps us calculate the probability of type II error β. The normal distribution on the left in Figure 9.3 is the distribution of the test statistic $\overline{X}$ when the null hypothesis H_0: $\mu = 50$ is true (this is what is meant by the expression "under H_0: $\mu = 50$"), and the normal distribution on the right is the distribution of $\overline{X}$ when the alternative hypothesis is true and the value of the mean is 52 (or "under H_1: $\mu = 52$"). A type II error will be committed if the sample mean $\overline{X}$ falls between 48.5 and 51.5 (the critical region boundaries) when $\mu = 52$. As seen in Figure 9.3, this is just the probability that $48.5 \leq \overline{X} \leq 51.5$ when the true mean is $\mu = 52$, or the shaded area under the normal distribution centered at $\mu = 52$. Therefore, referring to Figure 9.3, we find that

$$\beta = P(48.5 \leq \overline{X} \leq 51.5 \text{ when } \mu = 52)$$

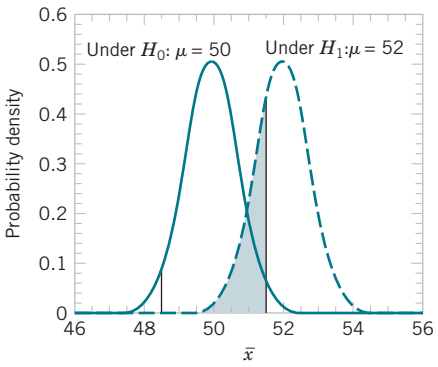


FIGURE 9.3

The probability of type II error when $\mu = 52$ and $n = 10$.

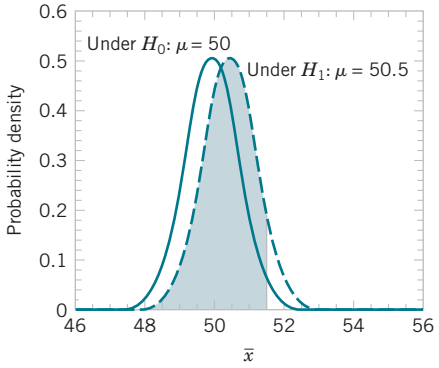


FIGURE 9.4

The probability of type II error when $\mu = 50.5$ and $n = 10$.

The z-values corresponding to 48.5 and 51.5 when $\mu = 52$ are

$$z_1 = \frac{48.5 - 52}{0.79} = -4.43 \qquad \text{and} \qquad z_2 = \frac{51.5 - 52}{0.79} = -0.63$$

Therefore,

$$\begin{aligned}\beta &= P(-4.43 \le Z \le -0.63) = P(Z \le -0.63) - P(Z \le -4.43)\\ &= 0.2643 - 0.0000 = 0.2643\end{aligned}$$

Thus, if we are testing H_0: $\mu = 50$ against H_1: $\mu \neq 50$ with $n = 10$ and the true value of the mean is $\mu = 52$, the probability that we will fail to reject the false null hypothesis is 0.2643. By symmetry, if the true value of the mean is $\mu = 48$, the value of β will also be 0.2643.

The probability of making a type II error β increases rapidly as the true value of μ approaches the hypothesized value. For example, see Figure 9.4, where the true value of the mean is $\mu = 50.5$ and the hypothesized value is H_0: $\mu = 50$. The true value of μ is very close to 50, and the value for β is

$$\beta = P(48.5 \le \overline{X} \le 51.5 \text{ when } \mu = 50.5)$$

As shown in Figure 9.4, the z-values corresponding to 48.5 and 51.5 when $\mu = 50.5$ are

$$z_1 = \frac{48.5 - 50.5}{0.79} = -2.53 \qquad \text{and} \qquad z_2 = \frac{51.5 - 50.5}{0.79} = 1.27$$

Therefore,

$$\begin{aligned}\beta &= P(-2.53 \le Z \le 1.27) = P(Z \le 1.27) - P(Z \le -2.53)\\ &= 0.8980 - 0.0057 = 0.8923\end{aligned}$$

Thus, the type II error probability is much higher for the case in which the true mean is 50.5 centimeters per second than for the case in which the mean is 52 centimeters per second. Of course, in many practical situations, we would not be as concerned with making a type II error if the mean were "close" to the hypothesized value. We would be much more interested in detecting large differences between the true mean and the value specified in the null hypothesis.

The type II error probability also depends on the sample size n. Suppose that the null hypothesis is H_0: $\mu = 50$ centimeters per second and that the true value of the mean is $\mu = 52$. If the sample size is increased from $n = 10$ to $n = 16$, the situation of Figure 9.5 results. The normal distribution on the left is the distribution of $\overline{X}$ when the mean $\mu = 50$, and the normal distribution

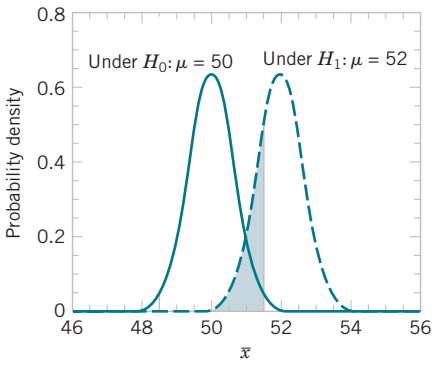


FIGURE 9.5

The probability of type II error when $\mu = 52$ and $n = 16$.

on the right is the distribution of $\overline{X}$ when $\mu = 52$. As shown in Figure 9.5, the type II error probability is

$$\beta = P(48.5 \le \overline{X} \le 51.5 \text{ when } \mu = 52)$$

When $n = 16$, the standard deviation of $\overline{X}$ is $\sigma/\sqrt{n} = 2.5/\sqrt{16} = 0.625$, and the z-values corresponding to 48.5 and 51.5 when $\mu = 52$ are

$$z_1 = \frac{48.5 - 52}{0.625} = -5.60 \qquad \text{and} \qquad z_2 = \frac{51.5 - 52}{0.625} = -0.80$$

Therefore,

$$\begin{aligned}\beta &= P(-5.60 \le Z \le -0.80) = P(Z \le -0.80) - P(Z \le -5.60)\\ &= 0.2119 - 0.0000 = 0.2119\end{aligned}$$

Recall that when $n = 10$ and $\mu = 52$, we found that $\beta = 0.2643$; therefore, increasing the sample size results in a decrease in the probability of type II error.

The results from this section and a few other similar calculations are summarized in the following table. The critical values are adjusted to maintain equal α for $n = 10$ and $n = 16$. This type of calculation is discussed later in the chapter.

Acceptance Region	Sample Size	α	β at $\mu = 52$	β at $\mu = 50.5$
$48.5 < \bar{x} < 51.5$	10	0.0576	0.2643	0.8923
$48 < \bar{x} < 52$	10	0.0114	0.5000	0.9705
$48.81 < \bar{x} < 51.19$	16	0.0576	0.0966	0.8606
$48.42 < \bar{x} < 51.58$	16	0.0114	0.2515	0.9578

The results in the boxes were not calculated in the text but the reader can easily verify them. This display and the discussion above reveal four important points:

1. The size of the critical region, and consequently the probability of a type I error α, can always be reduced by appropriate selection of the critical values.
2. Type I and type II errors are related. A decrease in the probability of one type of error always results in an increase in the probability of the other provided that the sample size n does not change.
3. An increase in sample size reduces β provided that α is held constant.
4. When the null hypothesis is false, β increases as the true value of the parameter approaches the value hypothesized in the null hypothesis. The value of β decreases as the difference between the true mean and the hypothesized value increases.

Generally, the analyst controls the type I error probability α when he or she selects the critical values. Thus, it is usually easy for the analyst to set the type I error probability at (or near) any desired value. Because the analyst can directly control the probability of wrongly rejecting H_0, we always think of rejection of the null hypothesis H_0 as a *strong conclusion*.

Because we can control the probability of making a type I error (or significance level), a logical question is what value should be used. The type I error probability is a measure of risk, specifically, the risk of concluding that the null hypothesis is false when it really is not. So, the value of α should be chosen to reflect the consequences (economic, social, etc.) of incorrectly rejecting the null hypothesis. Smaller values of α would reflect more serious consequences and larger values of α would be consistent with less severe consequences. This is often hard to do, so what has evolved in much of scientific and engineering practice is to use the value $\alpha = 0.05$ in most situations unless information is available that this is an inappropriate choice. In the rocket propellant problem with $n = 10$, this would correspond to critical values of 48.45 and 51.55.

A widely used procedure in hypothesis testing is to use a type I error or significance level of $\alpha = 0.05$. This value has evolved through experience and may not be appropriate for all situations.

On the other hand, the probability of type II error β is not a constant but depends on the true value of the parameter. It also depends on the sample size that we have selected. Because the type II error probability β is a function of both the sample size and the extent to which the null hypothesis H_0 is false, it is customary to think of the decision to accept H_0 as a *weak conclusion* unless we know that β is acceptably small. Therefore, rather than saying we "accept H_0," we prefer the terminology "fail to reject H_0." Failing to reject H_0 implies that we have not found sufficient evidence to reject H_0, that is, to make a strong statement. Failing to reject H_0 does not necessarily mean that there is a high probability that H_0 is true. It may simply mean that more data are required to reach a strong conclusion. This can have important implications for the formulation of hypotheses.

A useful analog exists between hypothesis testing and a jury trial. In a trial, the defendant is assumed innocent (this is like assuming the null hypothesis to be true). If strong evidence is found to the contrary, the defendant is declared to be guilty (we reject the null hypothesis). If evidence is insufficient, the defendant is declared to be not guilty. This is not the same as proving the defendant innocent and so, like failing to reject the null hypothesis, it is a weak conclusion.

An important concept that we use is the power of a statistical test.

Power

The **power** of a statistical test is the probability of rejecting the null hypothesis H_0 when the alternative hypothesis is true.

The power is computed as $1 - \beta$, and power can be interpreted as *the probability of correctly rejecting a false null hypothesis*. We often compare statistical tests by comparing their power properties. For example, consider the propellant burning rate problem when we are testing H_0: $\mu = 50$ centimeters per second against H_1: $\mu \neq 50$ centimeters per second. Suppose that the true value of the mean is $\mu = 52$. When $n = 10$, we found that $\beta = 0.2643$, so the power of this test is $1 - \beta = 1 - 0.2643 = 0.7357$ when $\mu = 52$.

Power is a very descriptive and concise measure of the *sensitivity* of a statistical test when by sensitivity we mean the ability of the test to detect differences. In this case, the sensitivity of the test for detecting the difference between a mean burning rate of 50 centimeters per second and

52 centimeters per second is 0.7357. That is, if the true mean is really 52 centimeters per second, this test will correctly reject H_0: $\mu = 50$ and "detect" this difference 73.57% of the time. If this value of power is judged to be too low, the analyst can increase either α or the sample size n.

9.1.3 One-Sided and Two-Sided Hypotheses

In constructing hypotheses, we always state the null hypothesis as an equality so that the probability of type I error α can be controlled at a specific value. The alternative hypothesis might be either one-sided or two-sided, depending on the conclusion to be drawn if H_0 is rejected. If the objective is to make a claim involving statements such as greater than, less than, superior to, exceeds, at least, and so forth, a one-sided alternative is appropriate. If no direction is implied by the claim, or if the claim "not equal to" is to be made, a two-sided alternative should be used.

EXAMPLE 9.1 | Propellant Burning Rate

Consider the propellant burning rate problem. Suppose that if the burning rate is less than 50 centimeters per second, we wish to show this with a strong conclusion. The hypotheses should be stated as

$$H_0: \mu = 50 \text{ centimeters per second}$$
$$H_1: \mu < 50 \text{ centimeters per second}$$

Here the critical region lies in the lower tail of the distribution of $\overline{X}$. Because the rejection of H_0 is always a strong conclusion, this statement of the hypotheses will produce the desired outcome if H_0 is rejected. Notice that, although the null hypothesis is stated with an equals sign, it is understood to include any value of μ not specified by the alternative hypothesis (that is, $\mu \leq 50$). Therefore, failing to reject H_0 does not mean that $\mu = 50$ centimeters per second exactly, but only that we do not have strong evidence in support of H_1.

In some real-world problems in which one-sided test procedures are indicated, selecting an appropriate formulation of the alternative hypothesis is occasionally difficult. For example, suppose that a soft-drink beverage bottler purchases 10-ounce bottles from a glass company. The bottler wants to be sure that the bottles meet the specification on mean internal pressure or bursting strength, which for 10-ounce bottles is a minimum strength of 200 psi. The bottler has decided to formulate the decision procedure for a specific lot of bottles as a hypothesis testing problem. There are two possible formulations for this problem, either

$$H_0: \mu = 200 \text{ psi} \qquad H_1: \mu > 200 \text{ psi} \tag{9.5}$$

or

$$H_0: \mu = 200 \text{ psi} \qquad H_1: \mu < 200 \text{ psi} \tag{9.6}$$

Consider the formulation in Equation 9.5. If the null hypothesis is rejected, the bottles will be judged satisfactory; if H_0 is not rejected, the implication is that the bottles do not conform to specifications and should not be used. Because rejecting H_0 is a strong conclusion, this formulation forces the bottle manufacturer to "demonstrate" that the mean bursting strength of the bottles exceeds the specification. Now consider the formulation in Equation 9.6. In this situation, the bottles will be judged satisfactory unless H_0 is rejected. That is, we conclude that the bottles are satisfactory unless there is strong evidence to the contrary.

Which formulation is correct, the one of Equation 9.5 or Equation 9.6? The answer is that it depends on the objective of the analysis. For Equation 9.5, there is some probability that H_0 will not be rejected (i.e., we would decide that the bottles are not satisfactory) even though the true mean is slightly greater than 200 psi. This formulation implies that we want the bottle manufacturer to demonstrate that the product meets or exceeds our specifications. Such a formulation could be appropriate if the manufacturer has experienced difficulty in meeting specifications in the past or if product safety considerations force us to hold tightly to the 200-psi specification. On the other hand, for the formulation of Equation 9.6, there is some probability that H_0 will

be accepted and the bottles judged satisfactory, even though the true mean is slightly less than 200 psi. We would conclude that the bottles are unsatisfactory only when there is strong evidence that the mean does not exceed 200 psi, that is, when H_0: $\mu = 200$ psi is rejected. This formulation assumes that we are relatively happy with the bottle manufacturer's past performance and that small deviations from the specification of $\mu \geq 200$ psi are not harmful.

In formulating one-sided alternative hypotheses, we should remember that rejecting H_0 is always a strong conclusion. Consequently, we should put the statement about which it is important to make a strong conclusion in the alternative hypothesis. In real-world problems, this will often depend on our point of view and experience with the situation.

9.1.4 *P*-Values in Hypothesis Tests

One way to report the results of a hypothesis test is to state that the null hypothesis was or was not rejected at a specified α-value or level of significance. This is called **fixed significance level** testing.

The fixed significance level approach to hypothesis testing is very nice because it leads directly to the concepts of type II error and power, which are of considerable value in determining the appropriate sample sizes to use in hypothesis testing. But the fixed significance level approach does have some disadvantages.

For example, in the propellant problem above, we can say that H_0: $\mu = 50$ was rejected at the 0.05 level of significance. This statement of conclusions may be often inadequate because it gives the decision maker no idea about whether the computed value of the test statistic was just barely in the rejection region or whether it was very far into this region. Furthermore, stating the results this way imposes the predefined level of significance on other users of the information. This approach may be unsatisfactory because some decision makers might be uncomfortable with the risks implied by $\alpha = 0.05$.

To avoid these difficulties, the ***P*-value** approach has been adopted widely in practice. The *P*-value is the probability that the test statistic will take on a value that is at least as extreme as the observed value of the statistic when the null hypothesis H_0 is true. Thus, a *P*-value conveys much information about the weight of evidence against H_0, and so a decision maker can draw a conclusion at *any* specified level of significance. We now give a formal definition of a *P*-value.

_P_-Value

The ***P*-value** is the smallest level of significance that would lead to rejection of the null hypothesis H_0 with the given data.

It is customary to consider the test statistic (and the data) significant when the null hypothesis H_0 is rejected; therefore, we may think of the *P*-value as the smallest level α at which the data are significant. In other words, the *P*-value is the **observed significance level**. Once the *P*-value is known, the decision maker can determine how significant the data are without the data analyst formally imposing a preselected level of significance.

Consider the two-sided hypothesis test for burning rate

$$H_0: \mu = 50 \quad H_1: \mu \neq 50$$

with $n = 16$ and $\sigma = 2.5$. Suppose that the observed sample mean is $\bar{x} = 51.3$ centimeters per second. Figure 9.6 is a critical region for this test with the value of $\bar{x} = 51.3$ and the symmetric

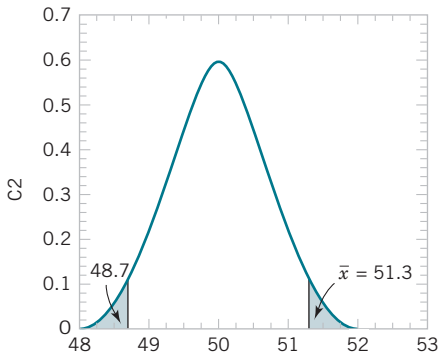


FIGURE 9.6

***P*-value is the area of the shaded region when $\bar{x} = 51.3$.**

value 48.7. The *P*-value of the test is the probability above 51.3 plus the probability below 48.7. The *P*-value is easy to compute after the test statistic is observed. In this example,

$$\begin{aligned} P\text{-value} &= 1 - P\left(48.7 < \overline{X} < 51.3\right) \\ &= 1 - P\left(\frac{48.7 - 50}{2.5/\sqrt{16}} < Z < \frac{51.3 - 50}{2.5/\sqrt{16}}\right) \\ &= 1 - P(-2.08 < Z < 2.08) \\ &= 1 - 0.962 = 0.038 \end{aligned}$$

The *P*-value tells us that if the null hypothesis $H_0 = 50$ is true, the probability of obtaining a random sample whose mean is at least as far from 50 as 51.3 (or 48.7) is 0.038. Therefore, an observed sample mean of 51.3 is a fairly rare event if the null hypothesis $H_0 = 50$ is really true. Compared to the "standard" level of significance 0.05, our observed *P*-value is smaller, so if we were using a fixed significance level of 0.05, the null hypothesis would be rejected. In fact, the null hypothesis $H_0 = 50$ would be rejected at *any* level of significance greater than or equal to 0.038. This illustrates the previous boxed definition; the *P*-value is the smallest level of significance that would lead to rejection of $H_0 = 50$.

Operationally, once a *P*-value is computed, we typically compare it to a predefined significance level to make a decision. Often this predefined significance level is 0.05. However, in presenting results and conclusions, it is standard practice to report the observed *P*-value along with the decision that is made regarding the null hypothesis.

Clearly, the *P*-value provides a measure of the credibility of the null hypothesis. Specifically, it is the risk that we have made an incorrect decision if we reject the null hypothesis H_0. The *P*-value is *not* the probability that the null hypothesis is false, nor is $1 - P$ the probability that the null hypothesis is true. The null hypothesis is either true or false (there is no probability associated with this), so the proper interpretation of the *P*-value is in terms of the risk of wrongly rejecting the null hypothesis H_0.

Computing the exact *P*-value for a statistical test is not always easy. However, most modern statistics software packages report the results of hypothesis testing problems in terms of *P*-values. We use the *P*-value approach extensively here.

More About *P*-Values We have observed that the procedure for testing a statistical hypothesis consists of drawing a random sample from the population, computing an appropriate statistic, and using the information in that statistic to make a decision regarding the null hypothesis. For example, we have used the sample average in decision making. Because the sample average is a random variable, its value will differ from sample to sample, meaning that the *P*-value associated with the test procedure is also a random variable. It also will differ from sample to sample. We are going to use a computer experiment (a simulation) to show how the *P*-value behaves when the null hypothesis is true and when it is false.

Consider testing the null hypothesis H_0: $\mu = 0$ against the alternative hypothesis H_0: $\mu \neq 0$ when we are sampling from a normal population with standard deviation $\sigma = 1$. Consider first the

case in which the null hypothesis is true and let's suppose that we are going to test the preceding hypotheses using a sample size of $n = 10$. We wrote a computer program to simulate drawing 10,000 different samples at random from a normal distribution with $\mu = 0$ and $\sigma = 1$. Then we calculated the *P*-values based on the values of the sample averages. Figure 9.7 is a histogram of the *P*-values obtained from the simulation. Notice that the histogram of the *P*-values is relatively uniform or flat over the interval from 0 to 1. It turns out that just slightly less than 5% of the *P*-values are in the interval from 0 to 0.05. It can be shown theoretically that if the null hypothesis is true, the probability distribution of the *P*-value is exactly uniform on the interval from 0 to 1. Because the null hypothesis is true in this situation, we have demonstrated by simulation that if a test of significance level 0.05 is used, the probability of wrongly rejecting the null hypothesis is (approximately) 0.05.

Now let's see what happens when the null hypothesis is false. We changed the mean of the normal distribution to $\mu = 1$ and repeated the previous computer simulation experiment by drawing another 10,000 samples and computing the *P*-values. Figure 9.8 is the histogram of the simulated *P*-values for this situation. Notice that this histogram looks very different from the one in Figure 9.7; there is a tendency for the *P*-values to stack up near the origin with many more small values between 0 and 0.05 than in the case in which the null hypothesis was true. Not all of the *P*-values are less than 0.05; those that exceed 0.05 represent type II errors or cases in which the null hypothesis is not rejected at the 0.05 level of significance even though the true mean is not 0.

Finally, Figure 9.9 shows the simulation results when the true value of the mean is even larger; in this case, $\mu = 2$. The simulated *P*-values are shifted even more toward 0 and concentrated on the left side of the histogram. Generally, as the true mean moves farther and farther away

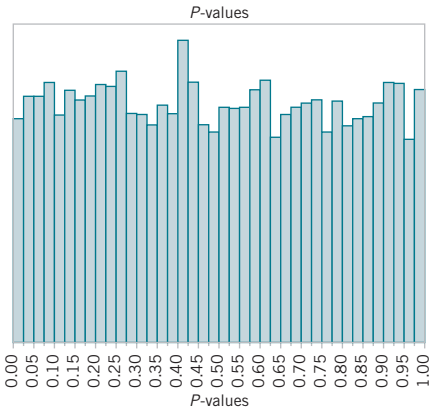


FIGURE 9.7

A *P*-value simulation when H_0: $\mu = 0$ is true.

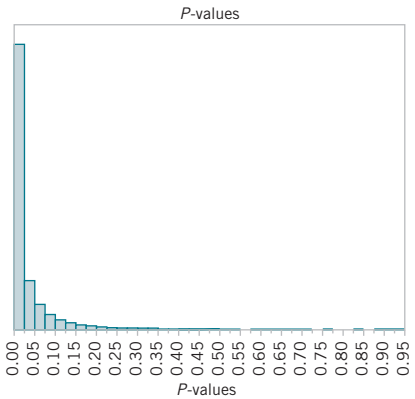


FIGURE 9.8

A *P*-value simulation when $\mu = 1$.

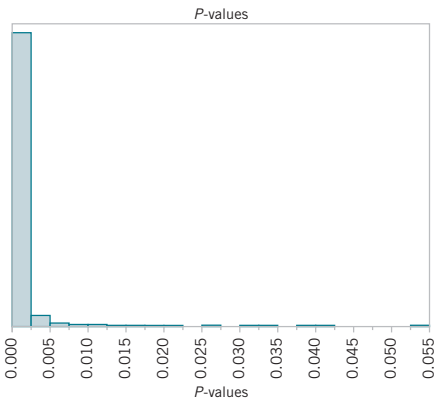


FIGURE 9.9

A *P*-value simulation when $\mu = 2$.

from the hypothesized value of 0 the distribution of the P-values will become more and more concentrated near 0 and fewer and fewer values will exceed 0.05. That is, the farther the mean is from the value specified in the null hypothesis, the higher is the chance that the test procedure will correctly reject the null hypothesis.

9.1.5 Connection between Hypothesis Tests and Confidence Intervals

A close relationship exists between the test of a hypothesis about any parameter, say θ, and the confidence interval for θ. If $[l, u]$ is a $100(1-\alpha)\%$ confidence interval for the parameter θ, the test with level of significance α of the hypothesis

$$H_0: \theta = \theta_0 \qquad H_1: \theta \neq \theta_0$$

will lead to rejection of H_0 if and only if θ_0 is *not* in the $100(1-\alpha\%)$ CI $[l, u]$. As an illustration, consider the escape system propellant problem with $\bar{x} = 51.3$, $\sigma = 2.5$, and $n = 16$. The null hypothesis H_0: $\mu = 50$ was rejected, using $\alpha = 0.05$. The 95% two-sided CI on μ can be calculated using Equation 8.7. This CI is $51.3 \pm 1.96(2.5/\sqrt{16})$ and this is $50.075 \leq \mu \leq 52.525$. Because the value $\mu_0 = 50$ is not included in this interval, the null hypothesis H_0: $\mu = 50$ is rejected.

Although hypothesis tests and CIs are equivalent procedures insofar as decision making or **inference** about μ is concerned, each provides somewhat different insights. For instance, the confidence interval provides a range of likely values for μ at a stated confidence level whereas hypothesis testing is an easy framework for displaying the *risk levels* such as the P-value associated with a specific decision. We continue to illustrate the connection between the two procedures throughout the text.

9.1.6 General Procedure for Hypothesis Tests

This chapter develops hypothesis-testing procedures for many practical problems. Use of the following sequence of steps in applying hypothesis-testing methodology is recommended.

1. **Parameter of interest:** From the problem context, identify the parameter of interest.
2. **Null hypothesis, H_0:** State the null hypothesis, H_0.
3. **Alternative hypothesis, H_1:** Specify an appropriate alternative hypothesis, H_1.
4. **Test statistic:** Determine an appropriate test statistic.
5. **Reject H_0 if:** State the rejection criteria for the null hypothesis.
6. **Computations:** Compute any necessary sample quantities, substitute these into the equation for the test statistic, and compute that value.
7. **Draw conclusions:** Decide whether or not H_0 should be rejected and report that in the problem context.

Steps 1–4 should be completed prior to examining the sample data. This sequence of steps is illustrated in subsequent sections.

In practice, such a formal and (seemingly) rigid procedure is not always necessary. Generally, once the experimenter (or decision maker) has decided on the question of interest and has determined the *design of the experiment* (that is, how the data are to be collected, how the measurements are to be made, and how many observations are required), only three steps are really required:

1. Specify the test statistic to be used (such as Z_0).
2. Specify the location of the critical region (two-tailed, upper-tailed, or lower-tailed).
3. Specify the criteria for rejection (typically, the value of α or the P-value at which rejection should occur).

These steps are often completed almost simultaneously in solving real-world problems, although we emphasize that it is important to think carefully about each step. That is why we present and use the seven-step process; it seems to reinforce the essentials of the correct approach. Although we may not use it every time in solving real problems, it is a helpful framework when we are first learning about hypothesis testing.

Statistical Versus Practical Significance We noted previously that reporting the results of a hypothesis test in terms of a P-value is very useful because it conveys more information than just the simple statement "reject H_0" or "fail to reject H_0." That is, rejection of H_0 at the 0.05 level of significance is much more meaningful if the value of the test statistic is well into the critical region, greatly exceeding the 5% critical value, than if it barely exceeds that value.

Even a very small P-value can be difficult to interpret from a practical viewpoint when we are making decisions because, although a small P-value indicates **statistical significance** in the sense that H_0 should be rejected in favor of H_1, the actual departure from H_0 that has been detected may have little (if any) **practical significance** (engineers like to say "engineering significance"). This is particularly true when the sample size n is large.

For example, consider the propellant burning rate problem in Example 9.1 in which we test H_0: $\mu = 50$ centimeters per second versus H_1: $\mu \neq 50$ centimeters per second with $\sigma = 2.5$. If we suppose that the mean rate is really 50.5 centimeters per second, this is not a serious departure from H_0: $\mu = 50$ centimeters per second in the sense that if the mean really is 50.5 centimeters per second, there is no practical observable effect on the performance of the air crew escape system. In other words, concluding that $\mu = 50$ centimeters per second when it is really 50.5 centimeters per second is an inexpensive error and has no practical significance. For a reasonably large sample size, a true value of $\mu = 50.5$ will lead to a sample $\bar{x}$ that is close to 50.5 centimeters per second, and we would not want this value of $\bar{x}$ from the sample to result in rejection of H_0. The following display shows the P-value for testing H_0: $\mu = 50$ when we observe $\bar{x} = 50.5$ centimeters per second and the power of the test at $\alpha = 0.05$ when the true mean is 50.5 for various sample sizes n:

Sample Size n	P-value When $\bar{x} = 50.5$	Power (at $\alpha = 0.05$) When True $\mu = 50.5$
10	0.527	0.097
25	0.317	0.170
50	0.157	0.293
100	0.046	0.516
400	6.3×10^{-5}	0.979
1000	2.5×10^{-10}	1.000

The P-value column in this display indicates that for large sample sizes, the observed sample value of $\bar{x} = 50.5$ would strongly suggest that H_0: $\mu = 50$ should be rejected, even though the observed sample results imply that from a practical viewpoint, the true mean does not differ much at all from the hypothesized value $\mu_0 = 50$. The power column indicates that if we test a hypothesis at a fixed significance level α, and even if there is little practical difference between the true mean and the hypothesized value, a large sample size will almost always lead to rejection of H_0. The moral of this demonstration is clear:

> Be careful when interpreting the results from hypothesis testing when the sample size is large because any small departure from the hypothesized value μ_0 will probably be detected, even when the difference is of little or no practical significance.

9.2 Tests on the Mean of a Normal Distribution, Variance Known

In this section, we consider hypothesis testing about the mean μ of a single normal population where the variance of the population σ^2 is known. We assume that a random sample $X_1, X_2, \ldots, X_n$ has been taken from the population. Based on our previous discussion, the sample mean $\overline{X}$ is an unbiased point estimator of μ with variance σ^2/n.

9.2.1 Hypothesis Tests on the Mean

Suppose that we wish to test the hypotheses

$$H_0: \mu = \mu_0 \qquad H_1: \mu \neq \mu_0 \tag{9.7}$$

where μ_0 is a specified constant. We have a random sample $X_1, X_2, \ldots, X_n$ from a normal population. Because $\overline{X}$ has a normal distribution (i.e., the *sampling distribution* of $\overline{X}$ is normal) with mean μ_0 and standard deviation $\sigma/\sqrt{n}$ if the null hypothesis is true, we could calculate a P-value or construct a critical region based on the computed value of the sample mean $\overline{X}$, as in Section 9.1.2.

It is usually more convenient to *standardize* the sample mean and use a test statistic based on the standard normal distribution. That is, the test procedure for $H_0: \mu = \mu_0$ uses the *test statistic*:

Test Statistic

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \tag{9.8}$$

If the null hypothesis $H_0: \mu = \mu_0$ is true, $E(\overline{X}) = \mu_0$, and it follows that the distribution of Z_0 is the standard normal distribution [denoted $N(0, 1)$].

The hypothesis testing procedure is as follows. Take a random sample of size n and compute the value of the sample mean $\overline{x}$. To test the null hypothesis using the P-value approach, we would find the probability of observing a value of the sample mean that is at least as extreme as $\overline{x}$, given that the null hypothesis is true. The standard normal z-value that corresponds to $\overline{x}$ is found from the test statistic in Equation 9.8:

$$z_0 = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

In terms of the standard normal cumulative distribution function (CDF), the probability we are seeking is $1 - \Phi(|z_0|)$. The reason that the argument of the standard normal cdf is $|z_0|$ is that the value of z_0 could be either positive or negative, depending on the observed sample mean. Because this is a two-tailed test, this is only one-half of the P-value. Therefore, for the two-sided alternative hypothesis, the P-value is

$$P = 2\left[1 - \Phi\left(|z_0|\right)\right] \tag{9.9}$$

This is illustrated in Figure 9.10(a).

Now let's consider the one-sided alternatives. Suppose that we are testing

$$H_0: \mu = \mu_0 \qquad H_1: \mu > \mu_0 \tag{9.10}$$

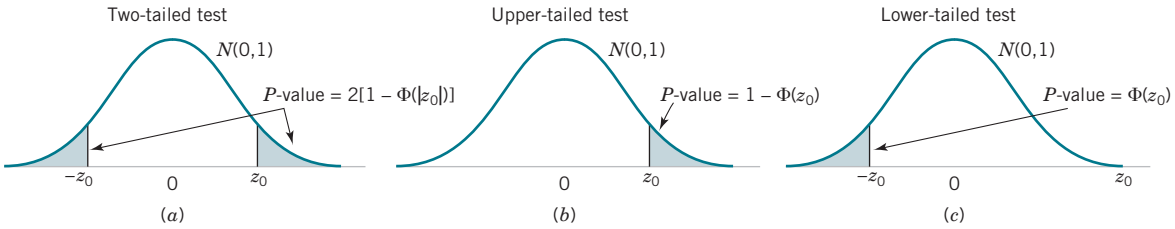


FIGURE 9.10

The *P*-value for a *z*-test. (a) The two-sided alternative H_1: $\mu \neq \mu_0$. (b) The one-sided alternative H_1: $\mu > \mu_0$. (c) The one-sided alternative H_1: $\mu < \mu_0$.

Once again, suppose that we have a random sample of size n and that the sample mean is $\bar{x}$. We compute the test statistic from Equation 9.8 and obtain z_0. Because the test is an upper-tailed test, only values of $\bar{x}$ that are greater than μ_0 are consistent with the alternative hypothesis. Therefore, the *P*-value would be the probability that the standard normal random variable is greater than the value of the test statistic z_0. This *P*-value is computed as

$$P = 1 - \Phi(z_0) \tag{9.11}$$

This *P*-value is shown in Figure 9.10(b).

The lower-tailed test involves the hypotheses

$$H_0: \mu = \mu_0 \qquad H_1: \mu < \mu_0 \tag{9.12}$$

Suppose that we have a random sample of size n and that the sample mean is $\bar{x}$. We compute the test statistic from Equation 9.8 and obtain z_0. Because the test is a lower-tailed test, only values of $\bar{x}$ that are less than μ_0 are consistent with the alternative hypothesis. Therefore, the *P*-value would be the probability that the standard normal random variable is less than the value of the test statistic z_0. This *P*-value is computed as

$$P = \Phi(z_0) \tag{9.13}$$

and shown in Figure 9.10(c). The **reference distribution** for this test is the standard normal distribution. The test is usually called a ***z*-test**.

We can also use the fixed significance level approach with the *z*-test. The only thing we have to do is determine where to place the critical regions for the two-sided and one-sided alternative hypotheses. First, consider the two-sided alternative in Equation 9.10. Now if H_0: $\mu = \mu_0$ is true, the probability is $1 - \alpha$ that the test statistic Z_0 falls between $-z_{\alpha/2}$ and $z_{\alpha/2}$ where $z_{\alpha/2}$ is the $100(\alpha/2)$ percentage point of the standard normal distribution. The regions associated with $z_{\alpha/2}$ and $-z_{\alpha/2}$ are illustrated in Figure 9.11(a). Note that the probability is α that the test statistic Z_0 will fall in the region $Z_0 > z_{\alpha/2}$ or $Z_0 < -z_{\alpha/2}$, when H_0: $\mu = \mu_0$ is true. Clearly, a sample producing

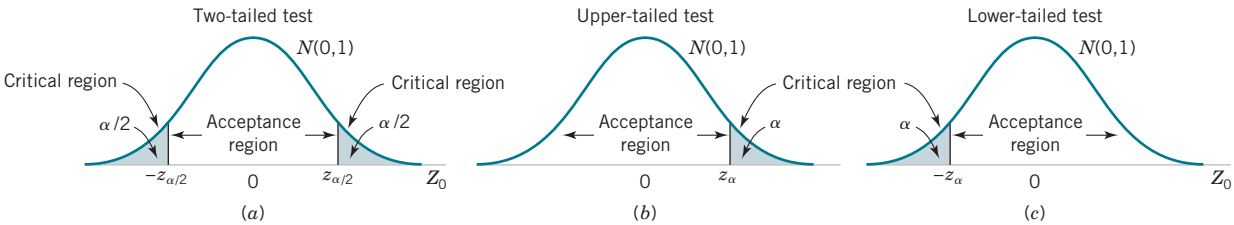


FIGURE 9.11

The distribution of Z_0 when H_0: $\mu = \mu_0$ is true with critical region for (a) the two-sided alternative H_1: $\mu \neq \mu_0$, (b) the one-sided alternative H_1: $\mu > \mu_0$, and (c) the one-sided alternative H_1: $\mu < \mu_0$.

a value of the test statistic that falls in the tails of the distribution of Z_0 would be unusual if H_0: $\mu = \mu_0$ is true; therefore, it is an indication that H_0 is false. Thus, we should reject H_0 if either

$$z_0 > z_{\alpha/2} \tag{9.14}$$

or

$$z_0 < -z_{\alpha/2} \tag{9.15}$$

and we should fail to reject H_0 if

$$-z_{\alpha/2} \le z_0 \le z_{\alpha/2} \tag{9.16}$$

Equations 9.14 and 9.15 define the **critical region** or **rejection region** for the test. The type I error probability for this test procedure is α.

We may also develop fixed significance level testing procedures for the one-sided alternatives. Consider the upper-tailed case in Equation 9.10.

In defining the critical region for this test, we observe that a negative value of the test statistic Z_0 would never lead us to conclude that H_0: $\mu = \mu_0$ is false. Therefore, we would place the critical region in the upper tail of the standard normal distribution and reject H_0 if the computed value z_0 is too large. Refer to Figure 9.11(b). That is, we would reject H_0 if

$$z_0 > z_\alpha \tag{9.17}$$

Similarly, to test the lower-tailed case in Equation 9.12, we would calculate the test statistic Z_0 and reject H_0 if the value of Z_0 is too small. That is, the critical region is in the lower tail of the standard normal distribution as in Figure 9.11(c), and we reject H_0 if

$$z_0 < -z_\alpha \tag{9.18}$$

Summary of Tests on the Mean, Variance Known

Testing Hypotheses on the Mean, Variance Known (*Z*-Tests)

Null hypothesis: H_0: $\mu = \mu_0$

Test statistic: $Z_0 = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$

Alternative Hypotheses	*P*-Value	Rejection Criterion for Fixed-Level Tests
H_1: $\mu \neq \mu_0$	Probability above $\lvert z_0 \rvert$ and probability below $-\lvert z_0 \rvert$, $P = 2[1 - \Phi(\lvert z_0 \rvert)]$	$z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$
H_1: $\mu > \mu_0$	Probability above z_0, $P = 1 - \Phi(z_0)$	$z_0 > z_\alpha$
H_1: $\mu < \mu_0$	Probability below z_0, $P = \Phi(z_0)$	$z_0 < -z_\alpha$

The *P*-values and critical regions for these situations are shown in Figures 9.10 and 9.11.

In general, understanding the critical region and the test procedure is easier when the test statistic is Z_0 rather than $\overline{X}$. However, the same critical region can always be written in terms of the computed value of the sample mean $\overline{x}$. A procedure identical to the preceding fixed significance level test is as follows:

$$\text{Reject } H_0\text{: } \mu = \mu_0 \text{ if either } \overline{x} > \alpha \text{ or } \overline{x} < b$$

where

$$a = \mu_0 + z_{\alpha/2}\sigma/\sqrt{n} \quad \text{and} \quad b = \mu_0 - z_{\alpha/2}\sigma/\sqrt{n}$$

EXAMPLE 9.2 | Propellant Burning Rate

Air crew escape systems are powered by a solid propellant. The burning rate of this propellant is an important product characteristic. Specifications require that the mean burning rate must be 50 centimeters per second. We know that the standard deviation of burning rate is $\sigma = 2$ centimeters per second. The experimenter decides to specify a type I error probability or significance level of $\alpha = 0.05$ and selects a random sample of $n = 25$ and obtains a sample average burning rate of $\bar{x} = 51.3$ centimeters per second. What conclusions should be drawn?

We may solve this problem by following the seven-step procedure outlined in Section 9.1.6. This results in

1. **Parameter of interest:** The parameter of interest is μ, the mean burning rate.
2. **Null hypothesis:** H_0: $\mu = 50$ centimeters per second
3. **Alternative hypothesis:** H_1: $\mu \neq 50$ centimeters per second
4. **Test statistic:** The test statistic is
$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$
5. **Reject H_0 if:** Reject H_0 if the P-value is less than 0.05. To use a fixed significance level test, the boundaries of the critical region would be $z_{0.025} = 1.96$ and $-z_{0.025} = -1.96$.
6. **Computations:** Because $\bar{x} = 51.3$ and $\sigma = 2$,
$$z_0 = \frac{51.3 - 50}{2/\sqrt{25}} = 3.25$$
7. **Conclusion:** Because the P-value $= 2[1 - \Phi(3.25)]$ $= 0.0012$ we reject H_0: $\mu = 50$ at the 0.05 level of significance.

Practical Interpretation: We conclude that the mean burning rate differs from 50 centimeters per second, based on a sample of 25 measurements. In fact, there is strong evidence that the mean burning rate exceeds 50 centimeters per second.

9.2.2 Type II Error and Choice of Sample Size

In testing hypotheses, the analyst directly selects the type I error probability. However, the probability of type II error β depends on the choice of sample size. In this section, we show how to calculate the probability of type II error β. We also show how to select the sample size to obtain a specified value of β.

Finding the Probability of Type II Error β

Consider the two-sided hypotheses

$$H_0\colon \mu = \mu_0 \qquad H_1\colon \mu \neq \mu_0$$

Suppose that the null hypothesis is false and that the true value of the mean is $\mu = \mu_0 + \delta$, say, where $\delta > 0$. The expected value of the test statistic Z_0 is

$$E(Z_0) = \frac{E(\bar{X}) - \mu_0}{\sigma/\sqrt{n}} = \frac{(\mu_0 + \delta) - \mu_0}{\sigma/\sqrt{n}} = \frac{\delta\sqrt{n}}{\sigma}$$

Therefore, the distribution of Z_0 when H_1 is true is

$$Z_0 \sim N\left(\frac{\delta\sqrt{n}}{\sigma}, 1\right) \tag{9.19}$$

The distribution of the test statistic Z_0 under both the null hypothesis H_0 and the alternate hypothesis H_1 is shown in Figure 9.9. From examining this figure, we note that if H_1 is true, a type II error will be made only if $-z_{\alpha/2} \leq Z_0 \leq z_{\alpha/2}$ where $Z_0 \sim N(\delta\sqrt{n}/\sigma, 1)$. That is, the probability of the type II error β is the probability that Z_0 falls between $-z_{\alpha/2}$ and $z_{\alpha/2}$ given that H_1 is true.

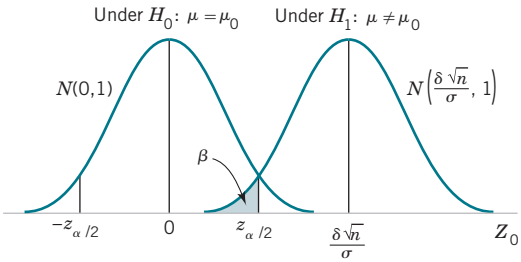


FIGURE 9.12
The distribution of Z_0 under H_0 and H_1.

This probability is shown as the shaded portion of Figure 9.12. Expressed mathematically, this probability is

Probability of a Type II Error for a Two-Sided Test on the Mean, Variance Known

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) \tag{9.20}$$

where $\Phi(z)$ denotes the probability to the left of z in the standard normal distribution. Note that Equation 9.20 was obtained by evaluating the probability that Z_0 falls in the interval $[-z_{\alpha/2}, z_{\alpha/2}]$ when H_1 is true. Furthermore, note that Equation 9.20 also holds if $\delta < 0$ because of the symmetry of the normal distribution. It is also possible to derive an equation similar to Equation 9.20 for a one-sided alternative hypothesis.

Sample Size Formulas One may easily obtain formulas that determine the appropriate sample size to obtain a particular value of β for a given Δ and α. For the two-sided alternative hypothesis, we know from Equation 9.20 that

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right)$$

or, if $\delta > 0$,

$$\beta \simeq \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) \tag{9.21}$$

because $\Phi(-z_{\alpha/2} - \delta\sqrt{n}/\sigma) \simeq 0$ when δ is positive. Let z_β be the 100β upper percentile of the standard normal distribution. Then, $\beta = \Phi(-z_\beta)$. From Equation 9.21,

$$-z_\beta \simeq z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}$$

or

Sample Size for a Two-Sided Test on the Mean, Variance Known

$$n \simeq \frac{(z_{\alpha/2} + z_\beta)^2\sigma^2}{\delta^2} \quad \text{where} \quad \delta = \mu - \mu_0 \tag{9.22}$$

If n is not an integer, the convention is to round the sample size up to the next integer. This approximation is good when $\Phi(-z_{\alpha/2} - \delta\sqrt{n}/\sigma)$ is small compared to β. For either of the one-sided alternative hypotheses, the sample size required to produce a specified type II error with probability β given δ and α is

Sample Size for a One-Sided Test on the Mean, Variance Known

$$n \simeq \frac{(z_\alpha + z_\beta)^2\sigma^2}{\delta^2} \quad \text{where} \quad \delta = \mu - \mu_0 \tag{9.23}$$

EXAMPLE 9.3 | Propellant Burning Rate Type II Error

Consider the rocket propellant problem in Example 9.2. Suppose that the true burning rate is 49 centimeters per second. What is β for the two-sided test with $\alpha = 0.05$, $\sigma = 2$, and $n = 25$?

Here $\delta = 1$ and $z_{\alpha/2} = 1.96$. From Equation 9.20,

$$\beta = \Phi\left(1.96 - \frac{\sqrt{25}}{\sigma}\right) - \Phi\left(-1.96 - \frac{\sqrt{25}}{\sigma}\right)$$
$$= \Phi(-0.54) - \Phi(-4.46) = 0.295$$

The probability is about 0.3 that this difference from 50 centimeters per second will not be detected. That is, the probability is about 0.3 that the test will fail to reject the null hypothesis when the true burning rate is 49 centimeters per second.

Practical Interpretation: A sample size of $n = 25$ results in reasonable, but not great, power $= 1 - \beta = 1 - 0.3 = 0.70$.

Suppose that the analyst wishes to design the test so that if the true mean burning rate differs from 50 centimeters per second by as much as 1 centimeter per second, the test will detect this (i.e., reject H_0: $\mu = 50$) with a high probability, say, 0.90. Now we note that $\sigma = 2$, $\delta = 51 - 50 = 1$, $\alpha = 0.05$, and $\beta = 0.10$. Because $z_{\alpha/2} = z_{0.025} = 1.96$ and $z_\beta = z_{0.10} = 1.28$, the sample size required to detect this departure from H_0: $\mu = 50$ is found by Equation 9.22 as

$$n \simeq \frac{(z_{\alpha/2} + z_\beta)^2\sigma^2}{\delta^2} = \frac{(1.96 + 1.28)^2 2^2}{(1^2)} \simeq 42$$

The approximation is good here because $\Phi(-z_{\alpha/2} - \delta\sqrt{n}/\sigma) = \Phi(-1.96 - (1)\sqrt{42}/2) = \Phi(-5.20) \simeq 0$, which is small relative to β.

Practical Interpretation: To achieve a much higher power of 0.90, you will need a considerably large sample size, $n = 42$ instead of $n = 25$.

Using Operating Characteristic Curves When performing sample size or type II error calculations, it is sometimes more convenient to use the **operating characteristic (OC) curves** in Appendix Charts VII*a* & *b*. These curves plot β as calculated from Equation 9.20 against a parameter d for various sample sizes n. Curves are provided for both $\alpha = 0.05$ and $\alpha = 0.01$. The parameter d is defined as

$$d = \frac{|\mu - \mu_0|}{\sigma} = \frac{|\delta|}{\sigma} \tag{9.24}$$

so one set of operating characteristic curves can be used for all problems regardless of the values of μ_0 and σ. From examining the operating characteristic curves or from Equation 9.20 and Figure 9.9, we note that

1. The farther the true value of the mean μ is from μ_0, the smaller the probability of type II error β for a given n and α. That is, we see that for a specified sample size and α, large differences in the mean are easier to detect than small ones.
2. For a given δ and α, the probability of type II error β decreases as n increases. That is, to detect a specified difference δ in the mean, we may make the test more powerful by increasing the sample size.

EXAMPLE 9.4 | Propellant Burning Rate Type II Error from OC Curve

Consider the propellant problem in Example 9.2. Suppose that the analyst is concerned about the probability of type II error if the true mean burning rate is $\mu = 51$ centimeters per second. We may use the operating characteristic curves to find β. Note that $\delta = 51 - 50 = 1$, $n = 25$, $\sigma = 2$, and $\alpha = 0.05$. Then using Equation 9.24 gives

$$d = \frac{|\mu - \mu_0|}{\sigma} = \frac{|\delta|}{\sigma} = \frac{1}{2}$$

and from Appendix Chart VII*a* with $n = 25$, we find that $\beta = 0.30$. That is, if the true mean burning rate is $\mu = 51$ centimeters per second, there is approximately a 30% chance that this will not be detected by the test with $n = 25$.

EXAMPLE 9.5 | Propellant Burning Rate Sample Size from OC Curve

Once again, consider the propellant problem in Example 9.2. Suppose that the analyst would like to design the test so that if the true mean burning rate differs from 50 centimeters per second by as much as 1 centimeter per second, the test will detect this (i.e., reject H_0: $\mu = 50$) with a high probability, say, 0.90. This is exactly the same requirement as in Example 9.3 in which we used Equation 9.22 to find the required sample size to be $n = 42$. The operating characteristic curves can also be used to find the sample size for this test. Because $d = |\mu - \mu_0|/\sigma = 1/2$, $\alpha = 0.05$, and $\beta = 0.10$, we find from Appendix Chart VII*a* that the required sample size is approximately $n = 40$. This closely agrees with the sample size calculated from Equation 9.22.

In general, the operating characteristic curves involve three parameters: β, d, and n. Given any two of these parameters, the value of the third can be determined. There are two typical applications of these curves:

1. For a given n and d, find β (as illustrated in Example 9.4). Analysts often encounter this kind of problem when they are concerned about the sensitivity of an experiment already performed, or when sample size is restricted by economic or other factors.
2. For a given β and d, find n. This was illustrated in Example 9.5. Analysts usually encounter this kind of problem when they have the opportunity to select the sample size at the outset of the experiment.

Operating characteristic curves are given in Appendix Charts VII*c* and VII*d* for the one-sided alternatives. If the alternative hypothesis is either H_1: $\mu > \mu_0$ or H_1: $\mu < \mu_0$, the abscissa scale on these charts is

$$d = \frac{|\mu - \mu_0|}{\sigma} \tag{9.25}$$

Using the Computer Many statistics software packages can calculate sample sizes and type II error probabilities. To illustrate, here are some typical computer calculations for the propellant burning rate problem:

Power and Sample Size

```
1-Sample Z-Test
Testing mean = null (versus not = null)
Calculating power for mean = null + difference
Alpha = 0.05 Sigma = 2
                    Sample            Target            Actual
Difference            Size             Power             Power
    1                   43            0.9000            0.9064
```

Power and Sample Size

```
1-Sample Z-Test
Testing mean = null (versus not = null)
Calculating power for mean = null + difference
Alpha = 0.05 Sigma = 2
                      Sample             Target             Actual
Difference             Size              Power              Power
    1                   28               0.7500             0.7536
```

Power and Sample Size

```
1-Sample Z-Test
Testing mean = null (versus not = null)
Calculating power for mean = null + difference
Alpha = 0.05 Sigma = 2
                      Sample
Difference             Size             Power
    1                   25              0.7054
```

In the first part of the boxed display, we worked Example 9.3, that is, to find the sample size n that would allow detection of a difference from $\mu_0 = 50$ of 1 centimeter per second with power of 0.9 and $\alpha = 0.05$. The answer, $n = 43$, agrees closely with the calculated value from Equation 9.22 in Example 9.3, which was $n = 42$. The difference is due to the software's use of a value of z_β that has more than two decimal places. The second part of the computer output relaxes the power requirement to 0.75. Note that the effect is to reduce the required sample size to $n = 28$. The third part of the output is the solution to Example 9.4 for which we wish to determine the type II error probability of (β) or the power $= 1 - \beta$ for the sample size $n = 25$. Note that software computes the power to be 0.7054, which agrees closely with the answer obtained from the OC curve in Example 9.4. Generally, however, the computer calculations will be more accurate than visually reading values from an OC curve.

9.2.3 Large-Sample Test

We have developed the test procedure for the null hypothesis H_0: $\mu = \mu_0$ assuming that the population is normally distributed and that σ^2 is known. In many if not most practical situations, σ^2 will be unknown. Furthermore, we may not be certain that the population is well modeled by a normal distribution. In these situations, if n is large (say, $n > 40$), the sample standard deviation s can be substituted for σ in the test procedures with little effect. Thus, although we have given a test for the mean of a normal distribution with known σ^2, it can be easily converted into a *large-sample test procedure for unknown* σ^2 that is valid regardless of the form of the distribution of the population. This large-sample test relies on the central limit theorem just as the large-sample confidence interval on μ that was presented in the previous chapter did. Exact treatment of the case in which the population is normal, σ^2 is unknown, and n is small involves use of the t distribution and is deferred until Section 9.3.

9.3 Tests on the Mean of a Normal Distribution, Variance Unknown

9.3.1 Hypothesis Tests on the Mean

We now consider the case of **hypothesis testing** on the mean of a population with *unknown variance* σ^2. The situation is analogous to the one in Section 8.2 where we considered a *confidence*

interval on the mean for the same situation. As in that section, the validity of the test procedure we describe rests on the assumption that the population distribution is at least approximately normal. The important result on which the test procedure relies is that if $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with mean μ and variance σ^2, the random variable

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has a t distribution with $n - 1$ degrees of freedom. Recall that we used this result in Section 8.2 to devise the t-confidence interval for μ. Now consider testing the hypotheses

$$H_0: \mu = \mu_0 \qquad H_1: \mu \neq \mu_0$$

We use the test statistic:

Test Statistic

$$T_0 = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \tag{9.26}$$

If the null hypothesis is true, T_0 has a t distribution with $n - 1$ degrees of freedom. When we know the distribution of the test statistic when H_0 is true (this is often called the **reference distribution** or the **null distribution**), we can calculate the P-value from this distribution, or, if we use a fixed significance level approach, we can locate the critical region to control the type I error probability at the desired level.

To test $H_0: \mu = \mu_0$ against the two-sided alternative $H_1: \mu \neq \mu_0$, the value of the test statistic t_0 in Equation 9.26 is calculated, and the P-value is found from the t distribution with $n - 1$ degrees of freedom (denoted by T_{n-1}). Because the test is two-tailed, the P-value is the sum of the probabilities in the two tails of the t distribution. Refer to Figure 9.13(a). The P-value is the probability above $|t_0|$ plus the probability below. Because the t distribution is symmetric around zero, a simple way to write this is

$$P = 2P(T_{n-1} > |t_0|) \tag{9.27}$$

A small P-value is evidence against H_0, so if P is of sufficiently small value (typically < 0.05), reject the null hypothesis.

For the one-sided alternative hypotheses,

$$H_0: \mu = \mu_0 \qquad H_1: \mu > \mu_0 \tag{9.28}$$

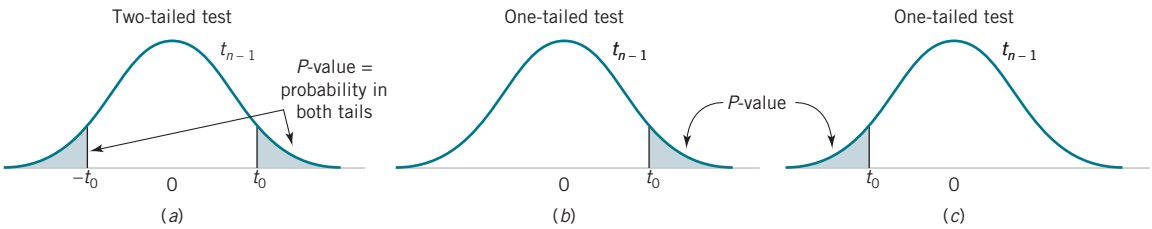


FIGURE 9.13

Calculating the P-value for a t-test: (a) $H_1: \mu \neq \mu_0$, (b) $H_1: \mu > \mu_0$, (c) $H_1: \mu < \mu_0$.

we calculate the test statistic t_0 from Equation 9.26 and calculate the P-value as

$$P = P(T_{n-1} > t_0) \tag{9.29}$$

For the other one-sided alternative,

$$H_0: \mu = \mu_0 \qquad H_1: \mu < \mu_0 \tag{9.30}$$

we calculate the P-value as

$$P = P(T_{n-1} < t_0) \tag{9.31}$$

Figure 9.13(b) and (c) show how these P-values are calculated.

Statistics software packages calculate and display P-values. However, in working problems by hand, it is useful to be able to find the P-value for a ***t*-test**. Because the t-table in Appendix A Table V contains only 10 critical values for each t distribution, determining the exact P-value from this table is usually impossible. Fortunately, it is easy to find lower and upper bounds on the P-value by using this table.

To illustrate, suppose that we are conducting an upper-tailed t-test (so $H_1: \mu > \mu_0$) with 14 degrees of freedom. The relevant critical values from Appendix A Table II are as follows:

Critical value:	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
Tail area:	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005

After calculating the test statistic, we find that $t_0 = 2.8$. Now $t_0 = 2.8$ is between two tabulated values, 2.624 and 2.977. Therefore, the P-value must be between 0.01 and 0.005. Refer to Figure 9.14. These are effectively the upper and lower bounds on the P-value.

This illustrates the procedure for an upper-tailed test. If the test is lower-tailed, just change the sign on the lower and upper bounds for t_0 and proceed in the same way. Remember that for a two-tailed test, the level of significance associated with a particular critical value is twice the corresponding tail area in the column heading. This consideration must be taken into account when we compute the bound on the P-value. For example, suppose that $t_0 = 2.8$ for a two-tailed alternative based on 14 degrees of freedom. The value of the test statistic $t_0 > 2.624$ (corresponding to $\alpha = 2 \times 0.01 = 0.02$) and $t_0 < 2.977$ (corresponding to $\alpha = 2 \times 0.005 = 0.01$), so the lower and upper bounds on the P-value would be $0.01 < P < 0.02$ for this case.

Some statistics software packages can calculate P-values. For example, many software packages have the capability to find cumulative probabilities from many standard probability

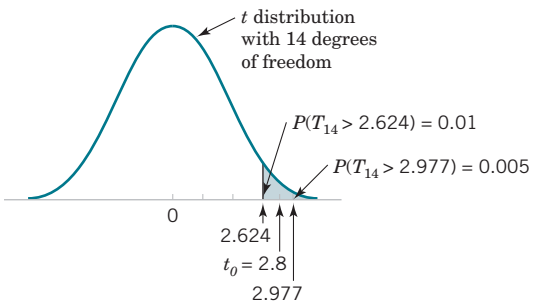


FIGURE 9.14

***P*-value for $t_0 = 2.8$; an upper-tailed test is shown to be between 0.005 and 0.01.**

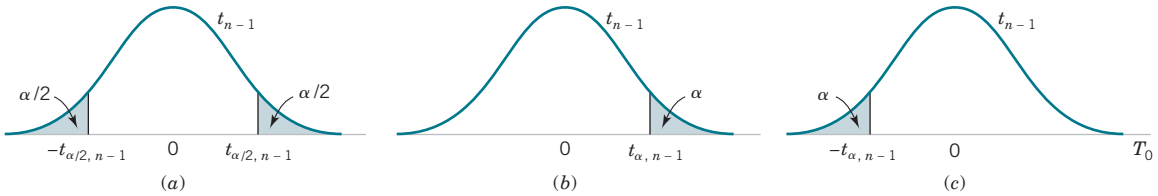


FIGURE 9.15

The distribution of T_0 when H_0: $\mu = \mu_0$ is true with critical region for (a) H_1: $\mu \neq \mu_0$, (b) H_1: $\mu > \mu_0$, and (c) H_1: $\mu < \mu_0$.

distributions, including the t distribution. Simply enter the value of the test statistic t_0 along with the appropriate number of degrees of freedom. Then the software will display the probability $P(T_\nu \leq t_o)$ where ν is the degrees of freedom for the test statistic t_0. From the cumulative probability, the P-value can be determined.

The single-sample t-test we have just described can also be conducted using the **fixed significance level** approach. Consider the two-sided alternative hypothesis. The null hypothesis would be rejected if the value of the test statistic t_0 falls in the critical region defined by the lower and upper $\alpha/2$ percentage points of the t distribution with $n - 1$ degrees of freedom. That is, reject H_0 if

$$t_0 > t_{\alpha/2,n-1} \quad \text{or} \quad t_0 < -t_{\alpha/2,n-1}$$

For the one-tailed tests, the location of the critical region is determined by the direction to which the inequality in the alternative hypothesis "points." So, if the alternative is H_1: $\mu > \mu_0$, reject H_0 if

$$t_0 > t_{\alpha,n-1}$$

and if the alternative is H_1: $\mu < \mu_0$, reject H_0 if

$$t_0 < -t_{\alpha,n-1}$$

Figure 9.15 provides the locations of these critical regions.

Summary for the One-Sample t-test

Testing Hypotheses on the Mean of a Normal Distribution, Variance Unknown

Null hypothesis: $H_0: \mu = \mu_0$

Test statistic: $T_0 = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$

Alternative Hypotheses	P-Value	Rejection Criterion for Fixed-Level Tests
$H_1: \mu \neq \mu_0$	Probability above $\lvert t_0 \rvert$ and probability below $-\lvert t_0 \rvert$	$t_0 > t_{\alpha/2,n-1}$ or $t_0 < -t_{\alpha/2,n-1}$
$H_1: \mu > \mu_0$	Probability above t_0	$t_0 > t_{\alpha,n-1}$
$H_1: \mu < \mu_0$	Probability below t_0	$t_0 < -t_{\alpha,n-1}$

The calculations of the P-values and the locations of the critical regions for these situations are shown in Figures 9.13 and 9.15, respectively.

EXAMPLE 9.6 | Golf Club Design

The increased availability of light materials with high strength has revolutionized the design and manufacture of golf clubs, particularly drivers. Clubs with hollow heads and very thin faces can result in much longer tee shots, especially for players of modest skills. This is due partly to the "spring-like effect" that the thin face imparts to the ball. Firing a golf ball at the head of the club and measuring the ratio of the ball's outgoing velocity to the incoming velocity can quantify this spring-like effect. The ratio of velocities is called the *coefficient of restitution of the club*. An experiment was performed in which 15 drivers produced by a particular club maker were selected at random and their coefficients of restitution measured. In the experiment, the golf balls were fired from an air cannon so that the incoming velocity and spin rate of the ball could be precisely controlled. It is of interest to determine whether there is evidence (with $\alpha = 0.05$) to support a claim that the mean coefficient of restitution exceeds 0.82. The observations follow:

0.8411	0.8191	0.8182	0.8125	0.8750
0.8580	0.8532	0.8483	0.8276	0.7983
0.8042	0.8730	0.8282	0.8359	0.8660

The sample mean and sample standard deviation are $\bar{x} = 0.83725$ and $s = 0.02456$. The normal probability plot of the data in Figure 9.16 supports the assumption that the coefficient of restitution is normally distributed. Because the experiment's objective is to demonstrate that the mean coefficient of restitution exceeds 0.82, a one-sided alternative hypothesis is appropriate.

The solution using the seven-step procedure for hypothesis testing is as follows:

1. **Parameter of interest:** The parameter of interest is the mean coefficient of restitution, μ.
2. **Null hypothesis:** H_0: $\mu = 0.82$
3. **Alternative hypothesis:** H_1: $\mu > 0.82$. We want to reject H_0 if the mean coefficient of restitution exceeds 0.82.
4. **Test statistic:** The test statistic is

$$t_0 = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

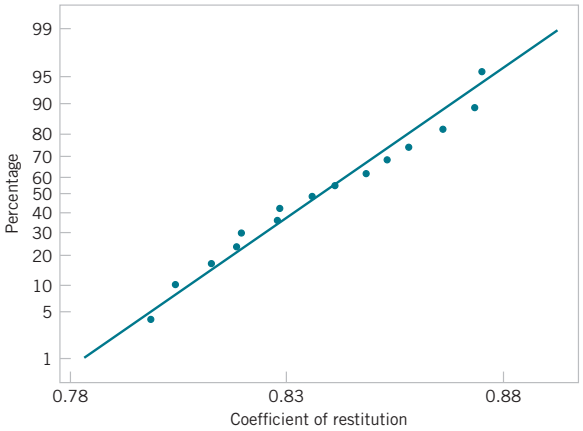


FIGURE 9.16

Normal probability plot of the coefficient of restitution data.

5. **Reject H_0 if:** Reject H_0 if the P-value is less than 0.05.
6. **Computations:** Because $\bar{x} = 0.83725$, $s = 0.02456$, $\mu_0 = 0.82$, and $n = 15$, we have

$$t_0 = \frac{0.83725 - 0.82}{0.02456/\sqrt{15}} = 2.72$$

7. **Conclusions:** From Appendix A Table II we find for a t distribution with 14 degrees of freedom that $t_0 = 2.72$ falls between two values: 2.624, for which $\alpha = 0.01$, and 2.977, for which $\alpha = 0.005$. Because this is a one-tailed test, we know that the P-value is between those two values, that is, $0.005 < P < 0.01$. Therefore, because $P < 0.05$, we reject H_0 and conclude that the mean coefficient of restitution exceeds 0.82.

Practical Interpretation: There is strong evidence to conclude that the mean coefficient of restitution exceeds 0.82.

Normality and the *t*-Test The development of the t-test assumes that the population from which the random sample is drawn is normal. This assumption is required to formally derive the t distribution as the reference distribution for the test statistic in Equation 9.26. Because it can be difficult to identify the form of a distribution based on a small sample, a logical question to ask is how important this assumption is. Studies have investigated this. Fortunately, studies have found that the t-test is relatively insensitive to the normality assumption. If the underlying population is reasonably symmetric and unimodal, the t-test will work satisfactorily. The exact significance level will not match the "advertised" level; for instance, the results may be significant at the 6%

or 7% level instead of the 5% level. This is usually not a serious problem in practice. A normal probability plot of the sample data as illustrated for the golf club data in Figure 9.16 is usually a good way to verify the adequacy of the normality assumption. Only severe departures from normality that are evident in the plot should be a cause for concern.

Many software packages conduct the one-sample t-test. Typical computer output for Example 9.6 is shown in the following display:

One-Sample T

Test of mu = 0.82 vs mu > 0.82

Variable	N	Mean	StDev	SE Mean	95.0% Lower confidence bound	T	P-value
COR	15	0.83725	0.02456	0.00634	0.82608	2.72	0.008

Notice that the software computes both the test statistic T_0 and a 95% lower confidence bound for the coefficient of restitution. The reported P-value is 0.008. Because the 95% lower confidence bound exceeds 0.82, we would reject the hypothesis that H_0: $\mu = 0.82$ and conclude that the alternative hypothesis H_1: $\mu > 0.82$ is true.

9.3.2 Type II Error and Choice of Sample Size

The type II error probability for the t-test depends on the distribution of the test statistic in Equation 9.26 when the null hypothesis H_0: $\mu = \mu_0$ is false. When the true value of the mean is $\mu = \mu_0 + \delta$, the distribution for T_0 is called the noncentral t distribution with $n - 1$ degrees of freedom and noncentrality parameter $\delta\sqrt{n}/\sigma$. Note that if $\delta = 0$, the noncentral t distribution reduces to the usual *central* t *distribution*. Therefore, the type II error of the two-sided alternative (for example) would be

$$\begin{aligned}\beta &= P(-t_{\alpha/2,n-1} \le T_0 \le t_{\alpha/2,n-1} \mid \delta \ne 0)\\ &= P(-t_{\alpha/2,n-1} \le T_0' \le t_{\alpha/2,n-1})\end{aligned}$$

where T_0' denotes the noncentral t random variable. Finding the type II error probability β for the t-test involves finding the probability contained between two points of the noncentral t distribution. Because the noncentral t-random variable has a messy density function, this integration must be done numerically.

Fortunately, this ugly task has already been done, and the results are summarized in a series of O.C. curves in Appendix Charts VII*e*, VII*f*, VII*g*, and VII*h* that plot β for the t-test against a parameter d for various sample sizes n. Curves are provided for two-sided alternatives on Charts VII*e* and VII*f*. The abscissa scale factor d on these charts is defined as

$$d = \frac{|\mu - \mu_0|}{\sigma} = \frac{|\delta|}{\sigma} \tag{9.32}$$

For the one-sided alternative $\mu > \mu_0$ or $\mu < \mu_0$, we use charts VII*g* and VII*h* with

$$d = \frac{|\mu - \mu_0|}{\sigma} = \frac{|\delta|}{\sigma} \tag{9.33}$$

We note that d depends on the unknown parameter σ^2. We can avoid this difficulty in several ways. In some cases, we may use the results of a previous experiment or prior information to make a rough initial estimate of σ^2. If we are interested in evaluating test performance after the data have been collected, we could use the sample variance s^2 to estimate σ^2. If there is no previous experience on which to draw in estimating σ^2, we then define the difference in the mean d that we wish to detect relative to σ. For example, if we wish to detect a small difference in the mean, we

might use a value of $d = |\delta|/\sigma \leq 1$ (for example), whereas if we are interested in detecting only moderately large differences in the mean, we might select $d = |\delta|/\sigma = 2$ (for example). That is, the value of the ratio $|\delta|/\sigma$ is important in determining sample size, and if it is possible to specify the relative size of the difference in means that we are interested in detecting, then a proper value of d can usually be selected.

EXAMPLE 9.7 | Golf Club Design Sample Size

Consider the golf club testing problem from Example 9.6. If the mean coefficient of restitution exceeds 0.82 by as much as 0.02, is the sample size $n = 15$ adequate to ensure that H_0: $\mu = 0.82$ will be rejected with probability at least 0.8?

To solve this problem, we use the sample standard deviation $s = 0.02456$ to estimate σ. Then $d = |\delta|/\sigma = 0.02/0.02456 = 0.81$. By referring to the operating characteristic curves in Appendix Chart VII*g* (for $\alpha = 0.05$) with $d = 0.81$ and $n = 15$, we find that $\beta = 0.10$, approximately. Thus, the probability of rejecting H_0: $\mu = 0.82$ if the true mean exceeds this by 0.02 is approximately $1 - \beta = 1 - 0.10 = 0.90$, and we conclude that a sample size of $n = 15$ is adequate to provide the desired sensitivity.

Some software packages can also perform power and sample size computations for the one-sample t-test. Several calculations based on the golf club testing problem follow:

Power and Sample Size

```
1-Sample t-test
Testing mean = null (versus > null)
Calculating power for mean = null + difference
Alpha = 0.05 Sigma = 0.02456

Difference          Sample Size         Power
0.02                     15             0.9117
```

Power and Sample Size

```
1-Sample t-test
Testing mean = null (versus > null)
Calculating power for mean = null + difference
Alpha = 0.05 Sigma = 0.02456

Difference          Sample Size         Power
0.01                     15             0.4425
```

Power and Sample Size

```
1-Sample t-test
Testing mean = null (versus > null)
Calculating power for mean = null + difference
Alpha = 0.05 Sigma = 0.02456

Difference          Sample Size          Target Power          Actual Power
0.01                     39                 0.8000                0.8029
```

In the first portion of the computer output, the software reproduces the solution to Example 9.7, verifying that a sample size of $n = 15$ is adequate to give power of at least 0.8 if the mean coefficient of restitution exceeds 0.82 by at least 0.02. In the middle section of the output, we used the software to compute the power to detect the difference between μ and $\mu_0 = 0.82$ of 0.01. Notice that with $n = 15$, the power drops considerably to 0.4425. The final portion of the output is the

sample size required for a power of at least 0.8 if the difference between μ and μ_0 of interest is actually 0.01. A much larger n is required to detect this smaller difference.

9.4 Tests on the Variance and Standard Deviation of a Normal Distribution

Sometimes hypothesis tests on the population variance or standard deviation are needed. When the population is modeled by a normal distribution, the tests and intervals described in this section are applicable.

9.4.1 Hypothesis Tests on the Variance

Suppose that we wish to test the hypothesis that the variance of a normal population σ^2 equals a specified value, say σ_0^2, or equivalently, that the standard deviation σ is equal to σ_0. Let $X_1, X_2, \ldots, X_n$ be a random sample of n observations from this population. To test

$$H_0: \sigma^2 = \sigma_0^2 \qquad H_1: \sigma^2 \neq \sigma_0^2 \tag{9.34}$$

we use the test statistic:

Test Statistic

$$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2} \tag{9.35}$$

If the null hypothesis $H_0: \sigma^2 = \sigma_0^2$ is true, the test statistic χ_0^2 defined in Equation 9.35 follows the chi-square distribution with $n-1$ degrees of freedom. This is the reference distribution for this test procedure. To perform a fixed significance level test, we would take a random sample from the population of interest, calculate χ_0^2, the value of the test statistic χ_0^2, and the null hypothesis $H_0: \sigma^2 = \sigma_0^2$ would be rejected if

$$\chi_0^2 > \chi_{\alpha/2,n-1}^2 \quad \text{or if} \quad \chi_0^2 > \chi_{1-\alpha/2,n-1}^2$$

where $\chi_{\alpha/2,n-1}^2$ and $\chi_{1-\alpha/2,n-1}^2$ are the upper and lower $100\alpha/2$ percentage points of the chi-square distribution with $n-1$ degrees of freedom, respectively. Figure 9.17(a) shows the critical region.

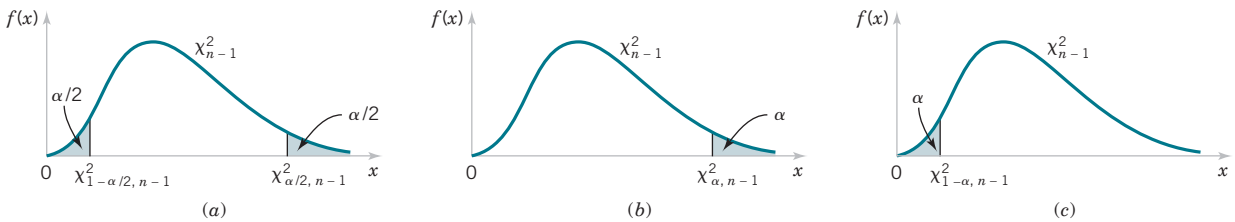


FIGURE 9.17

Reference distribution for the test of $H_0: \sigma^2 = \sigma_0^2$ with critical region values for (a) $H_1: \sigma^2 \neq \sigma_0^2$, (b) $H_1: \sigma^2 > \sigma_0^2$, and (c) $H_1: \sigma^2 < \sigma_0^2$.

The same test statistic is used for one-sided alternative hypotheses. For the one-sided hypotheses

$$H_0\colon \sigma^2 = \sigma_0^2 \qquad H_1\colon \sigma^2 > \sigma_0^2 \tag{9.36}$$

we would reject H_0 if $\chi_0^2 > \chi_{\alpha,n-1}^2$, whereas for the other one-sided hypotheses

$$H_0\colon \sigma^2 = \sigma_0^2 \qquad H_1\colon \sigma^2 < \sigma_0^2 \tag{9.37}$$

we would reject H_0 if $\chi_0^2 < \chi_{1-\alpha,n-1}^2$. The one-sided critical regions are shown in Figure 9.17(b) and (c).

Tests on the Variance of a Normal Distribution

Null hypothesis: $H_0\colon \sigma^2 = \sigma_0^2$

Test statistic: $\chi_0^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$

Alternative Hypothesis	Rejection Criteria
$H_1\colon \sigma^2 \neq \sigma_0^2$	$\chi_0^2 > \chi_{\alpha/2,n-1}^2$ or $\chi_0^2 < \chi_{1-\alpha/2,n-1}^2$
$H_1\colon \sigma^2 > \sigma_0^2$	$\chi_0^2 > \chi_{\alpha,n-1}^2$
$H_1\colon \sigma^2 < \sigma_0^2$	$\chi_0^2 < \chi_{1-\alpha,n-1}^2$

EXAMPLE 9.8 | Automated Filling

An automated filling machine is used to fill bottles with liquid detergent. A random sample of 20 bottles results in a sample variance of fill volume of $s^2 = 0.0153$ (fluid ounces)^2. If the variance of fill volume exceeds 0.01 (fluid ounces)^2, an unacceptable proportion of bottles will be underfilled or overfilled. Is there evidence in the sample data to suggest that the manufacturer has a problem with underfilled or overfilled bottles? Use $\alpha = 0.05$, and assume that fill volume has a normal distribution.

Using the seven-step procedure results in the following:

1. **Parameter of interest:** The parameter of interest is the population variance σ^2.
2. **Null hypothesis:** $H_0\colon \sigma^2 = 0.01$
3. **Alternative hypothesis:** $H_0\colon \sigma^2 > 0.01$
4. **Test statistic:** The test statistic is $\chi_0^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$
5. **Reject H_0 if:** Use $\alpha = 0.05$, and reject H_0 if $\chi_0^2 > \chi_{0.05,19}^2 = 30.14$
6. **Computations:** $\chi_0^2 = \dfrac{19(0.0153)}{0.01} = 29.07$
7. **Conclusions:** Because $\chi_0^2 = 29.07 < \chi_{0.05,19}^2 = 30.14$, we conclude that there is no strong evidence that the variance of fill volume exceeds 0.01 (fluid ounces)^2. So there is no strong evidence of a problem with incorrectly filled bottles.

We can also use the *P*-value approach. Using Appendix Table III, it is easy to place bounds on the *P*-value of a **chi-square test**. From inspection of the table, we find that $\chi_{0.10,19}^2 = 27.20$ and $\chi_{0.05,19}^2 = 30.14$. Because $27.20 < 29.07 < 30.14$, we conclude that the *P*-value for the test in Example 9.8 is in the interval $0.05 < P\text{-value} < 0.10$.

The *P*-value for a lower-tailed test would be found as the area (probability) in the lower tail of the chi-square distribution to the left of (or below) the computed value of the test statistic χ_0^2. For the two-sided alternative, find the tail area associated with the computed value of the test statistic and double it to obtain the *P*-value.

Some software packages perform the test on a variance of a normal distribution described in this section. Typical computer output for Example 9.8 is as follows:

Test and CI for One Variance

```
Null hypothesis                    Sigma-squared = 0.01
Alternative hypothesis             Sigma-squared > 0.01

Statistics
N                  StDev       Variance
20                 0.124       0 .0153

95% One-Sided Confidence Intervals

  Lower Confidence Bound          Lower Confidence Bound
  for StDev                            for Variance
  0.098                                 0.0096

Tests
Chi-Square            DF          P-Value
29.07                 19          0.065
```

Recall that we said that the *t*-test is relatively robust to the assumption that we are sampling from a normal distribution. The same is not true for the chi-square test on variance. Even moderate departures from normality can result in the test statistic in Equation 9.35 having a distribution that is very different from chi-square.

9.4.2 Type II Error and Choice of Sample Size

Operating characteristic curves for the chi-square tests in Section 9.4.1 are in Appendix Charts VI*i* through VI*n* for $\alpha = 0.05$ and $\alpha = 0.01$. For the two-sided alternative hypothesis of Equation 9.34, Charts VII*i* and VII*j* plot β against an abscissa parameter

$$\lambda = \frac{\sigma}{\sigma_0} \tag{9.38}$$

for various sample sizes n, where σ denotes the true value of the standard deviation. Charts VI*k* and VI*l* are for the one-sided alternative H_1: $\sigma^2 > \sigma_0^2$, and Charts VII*m* and VII*n* are for the other one-sided alternative H_1: $\sigma^2 < \sigma_0^2$. In using these charts, we think of σ as the value of the standard deviation that we want to detect.

These curves can be used to evaluate the β-error (or power) associated with a particular test. Alternatively, they can be used to *design* a test—that is, to determine what sample size is necessary to detect a particular value of σ that differs from the hypothesized value σ_0.

EXAMPLE 9.9 | Automated Filling Sample Size

Consider the bottle-filling problem from Example 9.8. If the variance of the filling process exceeds 0.01 (fluid ounces)^2, too many bottles will be underfilled. Thus, the hypothesized value of the standard deviation is $\sigma_0 = 0.10$. Suppose that if the true standard deviation of the filling process exceeds this value by 25%, we would like to detect this with probability at least 0.8. Is the sample size of $n = 20$ adequate?

To solve this problem, note that we require

$$\lambda = \frac{\sigma}{\sigma_0} = \frac{0.125}{0.10} = 1.25$$

This is the abscissa parameter for Chart VII*k*. From this chart, with $n = 20$ and $\lambda = 1.25$, we find that $\beta \simeq 0.6$. Therefore, there is only about a 40% chance that the null hypothesis will be rejected if the true standard deviation is really as large as $\sigma = 0.125$ fluid ounce.

To reduce the β-error, a larger sample size must be used. From the operating characteristic curve with $\beta = 0.20$ and $\lambda = 1.25$, we find that $n = 75$, approximately. Thus, if we want the test to perform as required, the sample size must be at least 75 bottles.

9.5 Tests on a Population Proportion

It is often necessary to test hypotheses on a population proportion. For example, suppose that a random sample of size n has been taken from a large (possibly infinite) population and that $X(\leq n)$ observations in this sample belong to a class of interest. Then $\widehat{P} = X/n$ is a point estimator of the proportion of the population p that belongs to this class. Note that n and p are the parameters of a binomial distribution. Furthermore, from Chapter 7, we know that the sampling distribution of $\widehat{P}$ is approximately normal with mean p and variance $p(1-p)/n$ if p is not too close to either 0 or 1 and if n is relatively large. Typically, to apply this approximation we require that np and $n(1-p)$ be greater than or equal to 5. We provide a large-sample test that uses the normal approximation to the binomial distribution.

9.5.1 Large-Sample Tests on a Proportion

Many engineering problems concern a random variable that follows the binomial distribution. For example, consider a production process that manufactures items that are classified as either acceptable or defective. Modelling the occurrence of defectives with the binomial distribution is usually reasonable when the binomial parameter p represents the proportion of defective items produced. Consequently, many engineering decision problems involve hypothesis testing about p.

We consider testing

$$H_0\colon p = p_0 \qquad H_1\colon p \neq p_0 \tag{9.39}$$

An approximate test based on the normal approximation to the binomial is given. As noted earlier, this approximate procedure will be valid as long as p is not extremely close to 0 or 1, and if the sample size is relatively large. Let X be the number of observations in a random sample of size n that belongs to the class associated with p. Then if the null hypothesis $H_0\colon p = p_0$ is true, we have $X \sim N[np_0, np_0(1-p_0)]$, approximately. To test $H_0\colon p = p_0$, calculate the test statistic

Test Statistic

$$Z_0 = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} \tag{9.40}$$

and determine the P-value. Because the test statistic follows a standard normal distribution if H_0 is true, the P-value is calculated exactly like the P-value for the z-tests in Section 9.2. So for the two-sided alternative hypothesis, the P-value is the sum of the probability in the standard normal distribution above $|z_0|$ and the probability below the negative value $-|z_0|$, or

$$P = 2\left[1 - \Phi(|z_0|)\right]$$

For the one-sided alternative hypothesis $H_1\colon p > p_0$, the P-value is the probability above z_0, or

$$P = 1 - \Phi(z_0)$$

and for the one-sided alternative hypothesis $H_1\colon p < p_0$, the P-value is the probability below z_0, or

$$P = \Phi(z_0)$$

We can also perform a fixed-significance-level test. For the two-sided alternative hypothesis, we would reject $H_0\colon p = p_0$ if

$$z_0 > z_{\alpha/2} \quad \text{or} \quad z_0 < -z_{\alpha/2}$$

Critical regions for the one-sided alternative hypotheses would be constructed in the usual manner.

Summary of Approximate Tests on a Binomial Proportion

Testing Hypotheses on a Binomial Proportion

Null hypotheses: $H_0: p = p_0$

Test statistic: $Z_0 = \dfrac{X - np_0}{\sqrt{np_0(1-p_0)}}$

Alternative Hypotheses	*P*-Value	Rejection Criterion for Fixed-Level Tests
$H_1: p \neq p_0$	Probability above $\lvert z_0\rvert$ and probability below $-\lvert z_0\rvert$, $P = 2[1 - \Phi(\lvert z_0\rvert)]$	$z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$
$H_1: p > p_0$	Probability above z_0, $P = 1 - \Phi(z_0)$	$z_0 > z_\alpha$
$H_1: p < p_0$	Probability below z_0, $P = \Phi(z_0)$	$z_0 < -z_\alpha$

EXAMPLE 9.10 | Automobile Engine Controller

A semiconductor manufacturer produces controllers used in automobile engine applications. The customer requires that the process fallout or fraction defective at a critical manufacturing step not exceed 0.05 and that the manufacturer demonstrate process capability at this level of quality using $\alpha = 0.05$. The semiconductor manufacturer takes a random sample of 200 devices and finds that four of them are defective. Can the manufacturer demonstrate process capability for the customer?

We may solve this problem using the seven-step hypothesis-testing procedure as follows:

1. **Parameter of interest:** The parameter of interest is the process fraction defective p.
2. **Null hypothesis:** $H_0: p = 0.05$
3. **Alternative hypothesis:** $H_1: p < 0.05$

   This formulation of the problem will allow the manufacturer to make a strong claim about process capability if the null hypothesis $H_0: p = 0.05$ is rejected.
4. **Test statistic:** The test statistic is (from Equation 9.40):

$$z_0 = \frac{x - np_0}{\sqrt{np_0(1-p_0)}}$$

   where $x = 4$, $n = 200$, and $p_0 = 0.05$.
5. **Reject H_0 if:** Reject $H_0: p = 0.05$ if the p-value is less than 0.05.
6. **Computation:** The test statistic is

$$z_0 = \frac{4 - 200(0.05)}{\sqrt{200(0.05)(0.95)}} = -1.95$$

7. **Conclusions:** Because $z_0 = -1.95$, the P-value is $\Phi(-1.95) = 0.0256$, so we reject H_0 and conclude that the process fraction defective p is less than 0.05.

Practical Interpretation: We conclude that the process is capable.

Another form of the test statistic Z_0 in Equation 9.40 is occasionally encountered. Note that if X is the number of observations in a random sample of size n that belongs to a class of interest, then $\hat{P} = X/n$ is the sample proportion that belongs to that class. Now divide both numerator and denominator of Z_0 in Equation 9.40 by n, giving

$$Z_0 = \frac{X/n - p_0}{\sqrt{p_0(1-p_0)/n}} \quad \text{or} \quad Z_0 = \frac{\hat{P} - p_0}{\sqrt{p_0(1-p_0)/n}} \tag{9.41}$$

This presents the test statistic in terms of the sample proportion instead of the number of items X in the sample that belongs to the class of interest.

Computer software packages can be used to perform the test on a binomial proportion. The following output shows typical results for Example 9.10.

Test and CI for One Proportion

Test of $p = 0.05$ vs $p < 0.05$

Sample	X	N	Sample p	95% Upper Confidence Bound	Z-Value	P-Value
1	4	200	0.020000	0.036283	−1.95	0.026

This output also shows a 95% one-sided upper-confidence bound on P. In Section 8.4, we showed how CIs on a binomial proportion are computed. This display shows the result of using the normal approximation for tests and CIs. When the sample size is small, this may be inappropriate.

Small Sample Tests on a Binomial Proportion Tests on a proportion when the sample size n is small are based on the binomial distribution, not the normal approximation to the binomial. To illustrate, suppose that we wish to test $H_0: p = p_0$ versus $H_0: p < p_0$. Let X be the number of successes in the sample. The P-value for this test would be found from the lower tail of a binomial distribution with parameters n and p_0. Specifically, the P-value would be the probability that a binomial random variable with parameters n and p_0 is less than or equal to X. P-values for the upper-tailed one-sided test and the two-sided alternative are computed similarly.

Many software packages calculate the exact P-value for a binomial test. The following output contains the exact P-value results for Example 9.10.

Test of $p = 0.05$ vs $p < 0.05$

Sample	X	N	Sample p	95% Upper Confidence Bound	Exact P-Value
1	4	200	0.020000	0.045180	0.026

The P-value is the same as that reported for the normal approximation because the sample size is fairly large. Notice that the CI is different from the one found using the normal approximation.

9.5.2 Type II Error and Choice of Sample Size

It is possible to obtain closed-form equations for the approximate β-error for the tests in Section 9.5.1. Suppose that p is the true value of the population proportion. The approximate β-error for the two-sided alternative $H_1: p \neq p_0$ is

$$\beta = \Phi\left(\frac{p_0 - p + z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) - \Phi\left(\frac{p_0 - p - z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) \tag{9.42}$$

If the alternative is $H_1: p < p_0$,

$$\beta = 1 - \Phi\left(\frac{p_0 - p - z_{\alpha}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) \tag{9.43}$$

whereas if the alternative is $H_1: p > p_0$,

$$\beta = \Phi\left(\frac{p_0 - p + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) \tag{9.44}$$

These equations can be solved to find the approximate sample size n that gives a test of level α that has a specified β risk. The sample size equations are

Approximate Sample Size for a Two-Sided Test on a Binomial Proportion

$$n = \left[\frac{z_{\alpha/2}\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p(1-p)}}{p - p_0}\right]^2 \tag{9.45}$$

for a two-sided alternative and for a one-sided alternative:

Approximate Sample Size for a One-Sided Test on a Binomial Proportion

$$n = \left[\frac{z_\alpha\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p(1-p)}}{p - p_0}\right]^2 \tag{9.46}$$

EXAMPLE 9.11 | Automobile Engine Controller Type II Error

Consider the semiconductor manufacturer from Example 9.10. Suppose that its process fallout is really $p = 0.03$. What is the β-error for a test of process capability that uses $n = 200$ and $\alpha = 0.05$?

The β-error can be computed using Equation 9.43 as follows:

$$\beta = 1 - \Phi\left[\frac{0.05 - 0.03 - (1.645)\sqrt{0.05(0.95)/200}}{\sqrt{0.03(1-0.03)/200}}\right]$$
$$= 1 - \Phi(-0.44) = 0.67$$

Thus, the probability is about 0.7 that the semiconductor manufacturer will fail to conclude that the process is capable if the true process fraction defective is $p = 0.03$ (3%). That is, the power of the test against this particular alternative is only about 0.3. This appears to be a large β-error (or small power), but the difference between $p = 0.05$ and $p = 0.03$ is fairly small, and the sample size $n = 200$ is not particularly large.

Suppose that the semiconductor manufacturer was willing to accept a β-error as large as 0.10 if the true value of the process fraction defective was $p = 0.03$. If the manufacturer continues to use $\alpha = 0.05$, what sample size would be required?

The required sample size can be computed from Equation 9.46 as follows:

$$n = \left[\frac{1.645\sqrt{0.05(0.95)} + 1.28\sqrt{0.03(0.97)}}{0.03 - 0.05}\right]^2 \simeq 832$$

where we have used $p = 0.03$ in Equation 9.46.

Conclusion: Note that $n = 832$ is a very large sample size. However, we are trying to detect a fairly small deviation from the null value $p_0 = 0.05$.

Some software packages also perform power and sample size calculations for the one-sample Z-test on a proportion. Typical computer output for the engine controllers tested in Example 9.10 follows.

Power and Sample Size

```
Test for One Proportion
Testing proportion = 0.05 (versus < 0.05)
Alpha = 0.05

Alternative   Sample
Proportion      Size   Power
3.00E-02         200   0.3287
```

Power and Sample Size

```
Test for One Proportion
Testing proportion = 0.05 (versus < 0.05)
Alpha = 0.05

Alternative   Sample    Target    Actual
Proportion      Size    Power     Power
3.00E-02         833    0.9000    0.9001
```

Power and Sample Size

```
Test for One Proportion
Testing proportion = 0.05 (versus < 0.05)
Alpha = 0.05

Alternative   Sample    Target    Actual
Proportion      Size    Power     Power
3.00E-02         561    0.7500    0.75030
```

The first part of the output shows the power calculation based on the situation described in Example 9.11 where the true proportion is really 0.03. The computer power calculation agrees with the results from Equation 9.43 in Example 9.11. The second part of the output computes the sample size necessary for a power of 0.9 ($\beta = 0.1$) if $p = 0.03$. Again, the results agree closely with those obtained from Equation 9.46. The final portion of the display shows the sample size that would be required if $p = 0.03$ and the power requirement is relaxed to 0.75. Notice that the sample size of $n = 561$ is still quite large because the difference between $p = 0.05$ and $p = 0.03$ is fairly small.

9.6 Summary Table of Inference Procedures for a Single Sample

The table in the end papers of this book (inside back cover) presents a summary of all the single-sample inference procedures from Chapters 8 and 9. The table contains the null hypothesis statement, the test statistic, the various alternative hypotheses and the criteria for rejecting H_0, and the formulas for constructing the $100(1 - \alpha)\%$ two-sided confidence interval. It would also be helpful to refer to the roadmap table in Chapter 8 that provides guidance to match the problem type to the information inside the back cover.

9.7 Testing for Goodness of Fit

The hypothesis-testing procedures that we have discussed in previous sections are designed for problems in which the population or probability distribution is known and the hypotheses involve the parameters of the distribution. Another kind of hypothesis is often encountered: We do not know the underlying distribution of the population, and we wish to test the hypothesis that a

particular distribution will be satisfactory as a population model. For example, we might wish to test the hypothesis that the population is normal.

We have previously discussed a very useful graphical technique for this problem called **probability plotting** and illustrated how it was applied in the case of a normal distribution. In this section, we describe a formal **goodness-of-fit test** procedure based on the chi-square distribution.

The test procedure requires a random sample of size n from the population whose probability distribution is unknown. These n observations are arranged in a frequency histogram, having k bins or class intervals. Let O_i be the observed frequency in the ith class interval. From the hypothesized probability distribution, we compute the expected frequency in the ith class interval, denoted E_i. The test statistic is

Goodness-of-Fit Test Statistic

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{9.47}$$

It can be shown that, if the population follows the hypothesized distribution, χ_0^2 has, approximately, a chi-square distribution with $k - p - 1$ degrees of freedom, when p represents the number of parameters of the hypothesized distribution estimated by sample statistics. This approximation improves as n increases. We should reject the null hypothesis that the population is the hypothesized distribution if the test statistic is too large. Therefore, the P-value would be the probability under the chi-square distribution with $k - p - 1$ degrees of freedom above the computed value of the test statistic χ_0^2 or $P = P(\chi_{k-p-1}^2 > \chi_0^2)$. For a fixed-level test, we would reject the hypothesis that the distribution of the population is the hypothesized distribution if the calculated value of the test statistic $\chi_0^2 > \chi_{\alpha,k-p-1}^2$.

One point to be noted in the application of this test procedure concerns the magnitude of the expected frequencies. If these expected frequencies are too small, the test statistic χ_0^2 will not reflect the departure of observed from expected but only the small magnitude of the expected frequencies. There is no general agreement regarding the minimum value of expected frequencies, but values of 3, 4, and 5 are widely used as minimal. Some writers suggest that an expected frequency could be as small as 1 or 2 so long as most of them exceed 5. Should an expected frequency be too small, it can be combined with the expected frequency in an adjacent class interval. The corresponding observed frequencies would then also be combined, and k would be reduced by 1. Class intervals are not required to be of equal width.

We now give two examples of the test procedure.

EXAMPLE 9.12 | Printed Circuit Board Defects—Poisson Distribution

The number of defects in printed circuit boards is hypothesized to follow a Poisson distribution. A random sample of $n = 60$ printed circuit boards has been collected, and the following number of defects observed.

Number of Defects	Observed Frequency
0	32
1	15
2	9
3	4

The mean of the assumed Poisson distribution in this example is unknown and must be estimated from the sample data. The estimate of the mean number of defects per board is the sample average, that is, $(32 \cdot 0 + 15 \cdot 1 + 9 \cdot 2 + 4 \cdot 3)/60 = 0.75$. From the Poisson distribution with parameter 0.75, we may compute p_i, the theoretical, hypothesized probability associated with the ith class interval. Because each class interval corresponds to a particular number of defects, we may find the p_i as follows:

$$p_1 = P(X = 0) = \frac{e^{-0.75}(0.75)^0}{0!} = 0.472$$

$$p_2 = P(X = 1) = \frac{e^{-0.75}(0.75)^1}{1!} = 0.354$$

$$p_3 = P(X = 2) = \frac{e^{-0.75}(0.75)^2}{2!} = 0.133$$

$$p_4 = P(X \geq 3) = 1 - (p_1 + p_2 + p_3) = 0.041$$

The expected frequencies are computed by multiplying the sample size $n = 60$ times the probabilities p_i. That is, $E_i = np_i$. The expected frequencies follow:

Number of Defects	Probability	Expected Frequency
0	0.472	28.32
1	0.354	21.24
2	0.133	7.98
3 (or more)	0.041	2.46

Because the expected frequency in the last cell is less than 3, we combine the last two cells:

Number of Defects	Observed Frequency	Expected Frequency
0	32	28.32
1	15	21.24
2 (or more)	13	10.44

The seven-step hypothesis-testing procedure may now be applied, using $\alpha = 0.05$, as follows:

1. **Parameter of interest:** The variable of interest is the form of the distribution of defects in printed circuit boards.
2. **Null hypothesis:** H_0: The form of the distribution of defects is Poisson.
3. **Alternative hypothesis:** H_1: The form of the distribution of defects is not Poisson.
4. **Test statistic:** The test statistic is $\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$
5. **Reject H_0 if:** Because the mean of the Poisson distribution was estimated, the preceding chi-square statistic will have $k - p - 1 = 3 - 1 - 1 = 1$ degree of freedom. Consider whether the P-value is less than 0.05.
6. **Computations:**

$$\chi_0^2 = \frac{(32 - 28.32)^2}{28.32} + \frac{(15 - 21.24)^2}{21.24} + \frac{(13 - 10.44)^2}{10.44} = 2.94$$

7. **Conclusions:** We find from Appendix Table III that $\chi^2_{0.10,1} = 2.71$ and $\chi^2_{0.05,1} = 3.84$. Because $\chi_0^2 = 2.94$ lies between these values, we conclude that the P-value is between 0.05 and 0.10. Therefore, because the P-value exceeds 0.05, we are unable to reject the null hypothesis that the distribution of defects in printed circuit boards is Poisson. The exact P-value computed from software is 0.0864.

EXAMPLE 9.13 | Power Supply Distribution—Continuous Distribution

A manufacturing engineer is testing a power supply used in a notebook computer and, using $\alpha = 0.05$, wishes to determine whether output voltage is adequately described by a normal distribution. Sample estimates of the mean and standard deviation of $\bar{x} = 5.04$ V and $s = 0.08$ V are obtained from a random sample of $n = 100$ units.

A common practice in constructing the class intervals for the frequency distribution used in the chi-square goodness-of-fit test is to choose the cell boundaries so that the expected frequencies $E_i = np_i$ are equal for all cells. To use this method, we want to choose the cell boundaries $a_0, a_1, \ldots, a_k$ for the k cells so that all the probabilities

$$p_i = P(a_{i-1} \le X \le a_i) = \int_{a_{i-1}}^{a_i} f(x)\,dx$$

are equal. Suppose that we decide to use $k = 8$ cells. For the standard normal distribution, the intervals that divide the scale into eight equally likely segments are (0, 0.32), (0.32, 0.675), (0.675, 1.15), (1.15, ∞), and their four "mirror image" intervals on the other side of zero. For each interval $p_i = 1/8 = 0.125$, so the expected cell frequencies are $E_i = n_{pi} = 100(0.125) = 12.5$. The complete table of observed and expected frequencies is as follows:

Class Interval	Observed Frequency o_i	Expected Frequency E_i
$x < 4.948$	12	12.5
$4.948 \le x < 4.986$	14	12.5
$4.986 \le x < 5.014$	12	12.5
$5.014 \le x < 5.040$	13	12.5
$5.040 \le x < 5.066$	12	12.5
$5.066 \le x < 5.094$	11	12.5
$5.094 \le x < 5.132$	12	12.5
$5.132 \le x$	14	12.5
Totals	100	100

The boundary of the first class interval is $\bar{x} - 1.15s = 4.948$. The second class interval is $[\bar{x} - 1.15s, \bar{x} - 0.675s]$ and so forth. We may apply the seven-step hypothesis-testing procedure to this problem.

1. **Parameter of interest:** The variable of interest is the form of the distribution of power supply voltage.
2. **Null hypothesis:** H_0: The form of the distribution is normal.

3. **Alternative hypothesis:** H_1: The form of the distribution is nonnormal.
4. **Test statistic:** The test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

5. **Reject H_0 if:** Because two parameters in the normal distribution have been estimated, the preceding chi-square statistic will have $k - p - 1 = 8 - 2 - 1 = 5$ degrees of freedom. We use a fixed significance level test with $\alpha = 0.05$. Therefore, we will reject H_0 if $\chi_0^2 > \chi_{0.05,5}^2 = 11.07$.
6. **Computations:**

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(o_i - E_i)^2}{E_i} = \frac{(12 - 12.5)^2}{12.5} + \frac{(14 - 12.5)^2}{12.5} + \cdots + \frac{(14 - 12.5)^2}{12.5} = 0.64$$

7. **Conclusions:** Because $\chi_0^2 = 0.64 < \chi_{0.05,5}^2 = 11.07$, we are unable to reject H_0, and no strong evidence indicates that output voltage is not normally distributed. The P-value for the chi-square statistic $\chi_0^2 = 0.64$ is $p = 0.9861$.

9.8 Contingency Table Tests

Many times the n elements of a sample from a population may be classified according to two different criteria. It is then of interest to know whether the two methods of classification are statistically **independent**; for example, we may consider the population of graduating engineers and may wish to determine whether starting salary is independent of academic disciplines. Assume that the first method of classification has r levels and that the second method has c levels. We will let O_{ij} be the observed frequency for level i of the first classification method and level j of the second classification method. The data would, in general, appear as shown in Table 9.2. Such a table is usually called an $r \times c$ **contingency table**.

We are interested in testing the hypothesis that the row-and-column methods of classification are independent. If we reject this hypothesis, we conclude some interaction exists between the two criteria of classification. The exact test procedures are difficult to obtain, but an approximate test statistic is valid for large n. Let p_{ij} be the probability that a randomly selected element falls in the ijth cell given that the two classifications are independent. Then $p_{ij} = u_i v_j$, where u_i is the probability that a randomly selected element falls in row class i and v_j is the probability that a randomly selected element falls in column class j. Now by assuming independence, the estimators of u_i and v_j are

$$\hat{u}_i = \frac{1}{n}\sum_{j=1}^{c} O_{ij} \qquad \hat{v}_j = \frac{1}{n}\sum_{i=1}^{r} O_{ij} \tag{9.48}$$

Therefore, the expected frequency of each cell is

$$E_{ij} = n\hat{u}_i\hat{v}_j = \frac{1}{n}\sum_{j=1}^{c} O_{ij}\sum_{i=1}^{r} O_{ij} \tag{9.49}$$

TABLE 9.2 An $r \times c$ Contingency Table

		Columns			
		1	2	⋯	c
	1	O_{11}	O_{12}	⋯	O_{1c}
Rows	2	O_{21}	O_{22}	⋯	O_{2c}
	⋮	⋮	⋮	⋮	⋮
	r	O_{r1}	O_{r2}	⋯	O_{rc}

Then, for large n, the statistic

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{9.50}$$

has an approximate chi-square distribution with $(r-1)(c-1)$ degrees of freedom if the null hypothesis is true. We should reject the null hypothesis if the value of the test statistic χ_0^2 is too large. The P-value would be calculated as the probability beyond χ_0^2 on the $\chi^2_{(r-1)(c-1)}$ distribution, or $P = P(\chi^2_{(r-1)(c-1)} > \chi_0^2)$. For a fixed-level test, we would reject the hypothesis of independence if the observed value of the test statistic χ_0^2 exceeded $\chi^2_{\alpha,(r-1)(c-1)}$.

EXAMPLE 9.14 | Health Insurance Plan Preference

A company has to choose among three health insurance plans. Management wishes to know whether the preference for plans is independent of job classification and wants to use $\alpha = 0.05$. The opinions of a random sample of 500 employees are shown in Table 9.3.

TABLE 9.3 Observed Data for Example 9.14

	Health Insurance Plan			
Job Classification	1	2	3	Totals
Salaried workers	160	140	40	340
Hourly workers	40	60	60	160
Totals	200	200	100	500

To find the expected frequencies, we must first compute $\hat{u}_1 = (340/500) = 0.68$, $\hat{u}_2 = (160/500) = 0.32$, $\hat{v}_1 = (200/500) = 0.40$, $\hat{v}_2 = (200/500) = 0.40$, and $\hat{v}_3 = (100/500) = 0.20$. The expected frequencies may now be computed from Equation 9.49. For example, the expected number of salaried workers favoring health insurance plan 1 is

$$E_{11} = n\hat{u}_1\hat{v}_1 = 500(0.68(0.40)) = 136$$

The expected frequencies are shown in Table 9.4.

TABLE 9.4 Expected Frequencies for Example 9.14

	Health Insurance Plan			
Job Classification	1	2	3	Totals
Salaried workers	136	136	68	340
Hourly workers	64	64	32	160
Totals	200	200	100	500

The seven-step hypothesis-testing procedure may now be applied to this problem.

1. **Parameter of interest:** The variable of interest is employee preference among health insurance plans.
2. **Null hypothesis:** H_0: Preference is independent of salaried versus hourly job classification.
3. **Alternative hypothesis:** H_1: Preference is not independent of salaried versus hourly job classification.
4. **Test statistic:** The test statistic is

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

5. **Reject H_0 if:** We will use a fixed-significance level test with $\alpha = 0.05$. Therefore, because $r = 2$ and $c = 3$, the degrees of freedom for chi-square are $(r-1)(c-1) = (1)(2) = 2$, and we would reject H_0 if $\chi_0^2 = \chi^2_{0.05,5} = 5.99$.
6. **Computations:**

$$\begin{aligned}\chi_0^2 &= \sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \\ &= \frac{(160-136)^2}{136} + \frac{(140-136)^2}{136} + \frac{(40-68)^2}{68} \\ &\quad + \frac{(40-64)^2}{64} + \frac{(60-64)^2}{64} + \frac{(60-32)^2}{32} \\ &= 49.63\end{aligned}$$

7. **Conclusions:** Because $\chi_0^2 = 49.63 > \chi^2_{0.05,2} = 5.99$, we reject the hypothesis of independence and conclude that the preference for health insurance plans is not independent of job classification. The P-value for $\chi_0^2 = 49.63$ is $P = 1.671 \times 10^{-11}$. (This value was computed by computer software.) Further analysis would be necessary to explore the nature of the association between these factors. It might be helpful to examine the table of observed minus expected frequencies.

Using the two-way contingency table to test independence between two variables of classification in a sample from a single population of interest is only one application of contingency table methods. Another common situation occurs when there are r populations of interest and each population is divided into the same c categories. A sample is then taken from the ith population, and the counts are entered in the appropriate columns of the ith row. In this situation, we want to investigate whether or not the proportions in the c categories are the same for all populations. The null hypothesis in this problem states that the populations are **homogeneous** with respect to the categories. For example, with only two categories, such as success and failure, defective and nondefective, and so on, the test for homogeneity is really a test of the equality of r binomial parameters. Calculation of expected frequencies, determination of degrees of freedom, and computation of the chi-square statistic for the test for homogeneity are identical to the test for independence.

9.9 Nonparametric Procedures

Most of the hypothesis-testing and confidence interval procedures discussed previously are based on the assumption that we are working with random samples from normal populations. Traditionally, we have called these procedures **parametric** methods because they are based on a particular parametric family of distributions—in this case, the normal. Alternately, sometimes we say that these procedures are not distribution free because they depend on the assumption of normality. Fortunately, most of these procedures are relatively insensitive to moderate departures from normality. In general, the t- and F-tests and the t-confidence intervals will have actual levels of significance or confidence levels that differ from the nominal or advertised levels chosen by the experimenter, although the difference in the actual and advertised levels is usually fairly small when the underlying population is not too different from the normal.

In this section, we describe procedures called *nonparametric* and *distribution-free methods*, and we usually make no assumptions about the distribution of the underlying population other than that it is continuous. These procedures have an accurate level of significance α or confidence level $100(1 - \alpha)\%$ for many different types of distributions. These procedures have some appeal. One of their advantages is that the data need not be quantitative but can be categorical (such as yes or no, defective or nondefective) or rank data. Another advantage is that nonparametric procedures are usually very quick and easy to perform.

The procedures described in this section are alternatives to the parametric t- and F-procedures described earlier. Consequently, it is important to compare the performance of both parametric and nonparametric methods under the assumptions of both normal and nonnormal populations. In general, nonparametric procedures do not utilize all the information provided by the sample. As a result, a nonparametric procedure will be less efficient than the corresponding parametric procedure when the underlying population is normal. This loss of efficiency is reflected by a requirement of a larger sample size for the nonparametric procedure than would be required by the parametric procedure to achieve the same power. On the other hand, this loss of efficiency is usually not large, and often the difference in sample size is very small. When the underlying distributions are not close to normal, nonparametric methods may have much to offer. They often provide improvement over the normal-theory parametric methods. Generally, if both parametric and nonparametric methods are applicable to a particular problem, we should use the more efficient parametric procedure.

Another approach is to **transform** the original data, say, by taking logarithms, square roots, or a reciprocal, and then analyze the transformed data using a parametric technique. A normal probability plot often works well to see whether the transformation has been successful. When this approach is successful, it is usually preferable to using a nonparametric technique. However, sometimes transformations are not satisfactory. That is, no transformation makes the sample observations look very close to a sample from a normal distribution. One situation in which this happens is when the data are in the form of **ranks**. These situations frequently occur in practice. For instance, a panel of judges may be used to evaluate 10 different formulations of a soft-drink beverage for overall quality with the "best" formulation assigned rank 1, the

"next-best" formulation assigned rank 2, and so forth. It is unlikely that rank data satisfy the normality assumption. Transformations may not prove satisfactory either. Many nonparametric methods involve the analysis of ranks and consequently are directly suited to this type of problem.

9.9.1 The Sign Test

The **sign test** is used to test hypotheses about the **median** $\tilde{\mu}$ of a continuous distribution. The median of a distribution is a value of the random variable X such that the probability is 0.5 that an observed value of X is less than or equal to the median, and the probability is 0.5 that an observed value of X is greater than or equal to the median. That is, $P(X \leq \tilde{\mu}) = P(X \geq \tilde{\mu}) = 0.5$.

Because the normal distribution is symmetric, the mean of a normal distribution equals the median. Therefore, the sign test can be used to test hypotheses about the mean of a normal distribution. This is the same problem for which we previously used the t-test. We briefly discuss the relative merits of the two procedures in Section 9.9.3. Note that, although the t-test was designed for samples from a normal distribution, the sign test is appropriate for samples from any continuous distribution. Thus, the sign test is a nonparametric procedure.

Suppose that the hypotheses are

$$H_0: \tilde{\mu} = \tilde{\mu}_0 \qquad H_1: \tilde{\mu} < \tilde{\mu}_0 \tag{9.51}$$

The test procedure is easy to describe. Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from the population of interest. Form the differences

$$X_i - \tilde{\mu}_0 \qquad i = 1, 2, \ldots, n \tag{9.52}$$

Now if the null hypothesis $H_0: \tilde{\mu} = \tilde{\mu}_0$ is true, any difference $X_i - \tilde{\mu}_0$ is equally likely to be positive or negative. An appropriate test statistic is the number of these differences that are positive, say, R^+. Therefore, to test the null hypothesis, we are really testing that the number of plus signs is a value of a binomial random variable that has the parameter $p = 1/2$. A P-value for the observed number of plus signs r^+ can be calculated directly from the binomial distribution. For instance, in testing the hypotheses in Equation 9.51, we will reject H_0 in favor of H_1 only if the proportion of plus signs is sufficiently less than 1/2 (or equivalently, when the observed number of plus signs r^+ is too small). Thus, if the computed P-value

$$P = P\left(R^+ \leq r^+ \text{ when } p = \frac{1}{2}\right)$$

is less than or equal to some preselected significance level α, we will reject H_0 and conclude that H_1 is true.

To test the other one-sided hypotheses

$$H_0: \tilde{\mu} = \tilde{\mu}_0 \qquad H_1: \tilde{\mu} > \tilde{\mu}_0 \tag{9.53}$$

we will reject H_0 in favor of H_1 only if the observed number of plus signs, say, r^+, is large or, equivalently, when the observed fraction of plus signs is significantly greater than 1/2. Thus, if the computed P-value

$$P = P\left(R^+ \geq r^+ \text{ when } p = \frac{1}{2}\right)$$

is less than α, we will reject H_0 and conclude that H_1 is true.

The two-sided alternative may also be tested. If the hypotheses are

$$H_0: \tilde{\mu} = \tilde{\mu}_0 \qquad H_1: \tilde{\mu} \neq \tilde{\mu}_0 \tag{9.54}$$

we should reject $H_0: \tilde{\mu} = \tilde{\mu}_0$ if the proportion of plus signs is significantly different from (either less than or greater than) 1/2. This is equivalent to the observed number of plus signs r^+ being either sufficiently large or sufficiently small. Thus, if $r^+ < n/2$, the P-value is

$$P = 2P\left(R^+ \leq r^+ \text{ when } p = \frac{1}{2}\right)$$

and if $r^+ > n/2$, the P-value is

$$P = 2P\left(R^+ \geq r^+ \text{ when } p = \frac{1}{2}\right)$$

If the P-value is less than some preselected level α, we will reject H_0 and conclude that H_1 is true.

EXAMPLE 9.15 | Propellant Shear Strength Sign Test

Montgomery, Peck, and Vining (2012) reported on a study in which a rocket motor is formed by binding an igniter propellant and a sustainer propellant together inside a metal housing. The shear strength of the bond between the two propellant types is an important characteristic. The results of testing 20 randomly selected motors are shown in Table 9.5. We would like to test the hypothesis that the median shear strength is 2000 psi, using $\alpha = 0.05$.

TABLE 9.5 Propellant Shear Strength Data

Observation i	Shear Strength x_i	Differences $x_i - 2000$	Sign
1	2158.70	+158.70	+
2	1678.15	−321.85	−
3	2316.00	+316.00	+
4	2061.30	+61.30	+
5	2207.50	+207.50	+
6	1708.30	−291.70	−
7	1784.70	−215.30	−
8	2575.10	+575.10	+
9	2357.90	+357.90	+
10	2256.70	+256.70	+
11	2165.20	+165.20	+
12	2399.55	+399.55	+
13	1779.80	−220.20	−
14	2336.75	+336.75	+
15	1765.30	−234.70	−
16	2053.50	+53.50	+
17	2414.40	+414.40	+
18	2200.50	+200.50	+
19	2654.20	+654.20	+
20	1753.70	−246.30	−

This problem can be solved using the seven-step hypothesis-testing procedure:

1. **Parameter of interest:** The parameter of interest is the median of the distribution of propellant shear strength.
2. **Null hypothesis:** H_0: $\tilde{\mu} = 2000$ psi
3. **Alternative hypothesis:** H_1: $\tilde{\mu} \neq 2000$ psi
4. **Test statistic:** The test statistic is the observed number of plus differences in Table 9.5, or $r^+ = 14$.
5. **Reject H_0 if:** We will reject H_0 if the P-value corresponding to $r^+ = 14$ is less than or equal to $\alpha = 0.05$.
6. **Computations:** Because $r^+ = 14$ is greater than $n/2 = 20/2 = 10$, we calculate the P-value from

$$P = 2P\left(R^+ \geq 14 \text{ when } p = \frac{1}{2}\right)$$
$$= 2\sum_{r=14}^{20} \binom{20}{r} (0.5)^r (0.5)^{20-r} = 0.1153$$

7. **Conclusions:** Because $p = 0.1153$ is not less than $\alpha = 0.05$, we cannot reject the null hypothesis that the median shear strength is 2000 psi. Another way to say this is that the observed number of plus signs $r^+ = 14$ was not large or small enough to indicate that median shear strength is different from 2000 psi at the $\alpha = 0.05$ level of significance.

It is also possible to construct a table of critical values for the sign test. This table is shown as Appendix Table VIII. Its use for the two-sided alternative hypothesis in Equation 9.54 is simple. As before, let R^+ denote the number of the differences $(X_i - \tilde{\mu}_0)$ that are positive and let R^-

denote the number of these differences that are negative. Let $R = \min(R^+, R^-)$. Appendix Table VIII presents critical values r_α^* for the sign test that ensure that P(type I error) = P(reject H_0 when H_0 is true) = α for $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.10$. If the observed value of the test statistic $r \le r_\alpha^*$, the null hypothesis H_0: $\tilde{\mu} = \tilde{\mu}_0$ should be rejected.

To illustrate how this table is used, refer to the data in Table 9.5 that were used in Example 9.15. Now $r^+ = 14$ and $r^- = 6$; therefore, $r = \min(14, 6) = 6$. From Appendix Table VIII with $n = 20$ and $\alpha = 0.05$, we find that $r_{0.05}^* = 5$. Because $r = 6$ is not less than or equal to the critical value $r_{0.05}^* = 5$, we cannot reject the null hypothesis that the median shear strength is 2000 psi.

We can also use Appendix Table VIII for the sign test when a one-sided alternative hypothesis is appropriate. If the alternative is H_1: $\tilde{\mu} > \tilde{\mu}_0$, reject H_0: $\tilde{\mu} = \tilde{\mu}_0$ if $r^- \le r_\alpha^*$; if the alternative is H_1: $\tilde{\mu} > \tilde{\mu}_0$, reject H_0: $\tilde{\mu} = \tilde{\mu}_0$ if $r^+ \le r_\alpha^*$. The level of significance of a one-sided test is one-half the value for a two-sided test. Appendix Table VIII shows the one-sided significance levels in the column headings immediately following the two-sided levels.

Finally, note that when a test statistic has a discrete distribution such as R does in the sign test, it may be impossible to choose a critical value r_α^* that has a level of significance exactly equal to α. The approach used in Appendix Table VIII is to choose r_α^* to yield an α that is as close to the advertised significance level α as possible.

Ties in the Sign Test Because the underlying population is assumed to be continuous, there is a zero probability that we will find a "tie"—that is, a value of X_i exactly equal to $\tilde{\mu}_0$. However, this may sometimes happen in practice because of the way the data are collected. When ties occur, they should be set aside and the sign test applied to the remaining data.

The Normal Approximation When $p = 0.5$, the binomial distribution is well approximated by a normal distribution when n is at least 10. Thus, because the mean of the binomial is np and the variance is $np(1 - p)$, the distribution of R^+ is approximately normal with mean $0.05n$ and variance $0.25n$ whenever n is moderately large. Therefore, in these cases, the null hypothesis H_0: $\tilde{\mu} = \tilde{\mu}_0$ can be tested using the statistic

Normal Approximation for Sign Test Statistic

$$Z_0 = \frac{R^+ - 0.5n}{0.5\sqrt{n}} \tag{9.55}$$

A P-value approach could be used for decision making. The fixed significance level approach could also be used.

The two-sided alternative would be rejected if the observed value of the test statistic $|z_0| > z_{\alpha/2}$, and the critical regions of the one-sided alternative would be chosen to reflect the sense of the alternative. (If the alternative is H_1: $\tilde{\mu} > \tilde{\mu}_0$, reject H_0 if $z_0 > z_\alpha$, for example.)

Type II Error for the Sign Test The sign test will control the probability of a type I error at an advertised level α for testing the null hypothesis H_0: $\tilde{\mu} = \tilde{\mu}_0$ for any continuous distribution. As with any hypothesis-testing procedure, it is important to investigate the probability of a type II error, β. The test should be able to effectively detect departures from the null hypothesis, and a good measure of this effectiveness is the value of β for departures that are important. A small value of β implies an effective test procedure.

In determining β, it is important to realize not only that a particular value of $\tilde{\mu}$, say, $\tilde{\mu}_0 + \Delta$, must be used but also that the **form** of the underlying distribution will affect the calculations. To illustrate, suppose that the underlying distribution is normal with $\sigma = 1$ and we are testing the hypothesis H_0: $\tilde{\mu} = 2$ versus H_1: $\tilde{\mu} > 2$. (Because $\tilde{\mu} = \mu$ in the normal distribution, this is equivalent to testing that the mean equals 2.) Suppose that it is important to detect a departure

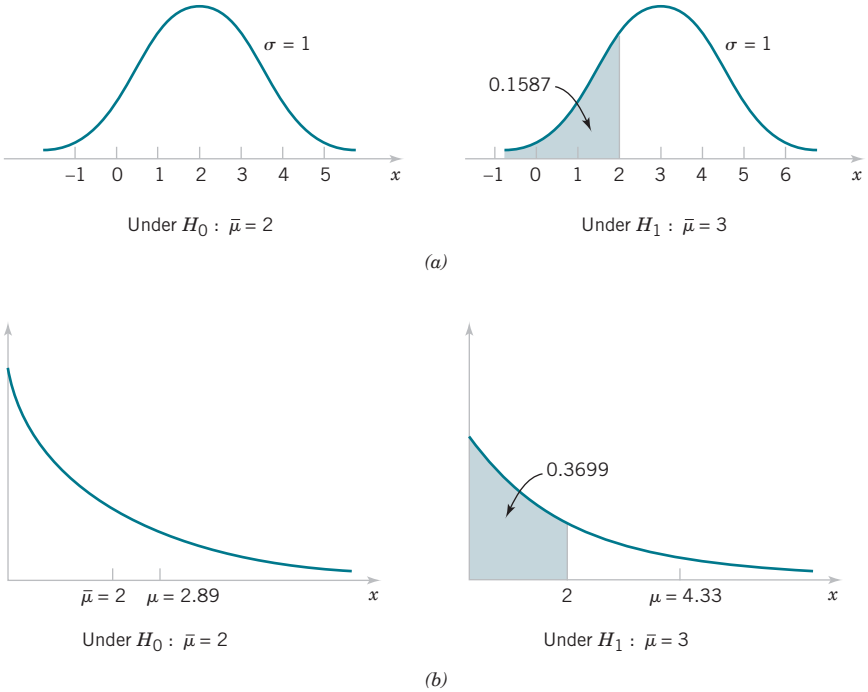

σ = 1
−1 0 1 2 3 4 5 x
Under H_0 : $\bar{\mu} = 2$
σ = 1
0.1587
−1 0 1 2 3 4 5 6 x
Under H_1 : $\bar{\mu} = 3$
(a)
$\bar{\mu} = 2$ $\mu = 2.89$ x
Under H_0 : $\bar{\mu} = 2$
0.3699
2 $\mu = 4.33$ x
Under H_1 : $\bar{\mu} = 3$
(b)

FIGURE 9.18

Calculation of β for the sign test. (a) Normal distributions. (b) Exponential distributions.

from $\tilde{\mu} = 2$ to $\tilde{\mu} = 3$. The situation is illustrated graphically in Figure 9.18(a). When the alternative hypothesis is true (H_1: $\tilde{\mu} = 3$), the probability that the random variable X is less than or equal to the value 2 is

$$P(X \leq 2) = P(Z \leq -1) = \Phi(-1) = 0.1587$$

Suppose that we have taken a random sample of size 12. At the $\alpha = 0.05$ level, Appendix Table VIII indicates that we would reject H_0: $\tilde{\mu} = 2$ if $r^- \leq r^*_{0.05} = 2$. Therefore, β is the probability that we do not reject H_0: μ when in fact $\tilde{\mu} = 3$, or

$$\beta = 1 - \sum_{x=0}^{2} \binom{12}{x} (0.1587)^x (0.8413)^{12-x} = 0.2944$$

If the distribution of X had been exponential rather than normal, the situation would be as shown in Figure 9.18(b), and the probability that the random variable X is less than or equal to the value $x = 2$ when $\tilde{\mu} = 3$ (note that when the median of an exponential distribution is 3, the mean is 4.33) is

$$P(X \leq 2) = \int_0^2 \frac{1}{4.33} e^{-\frac{1}{4.33}x} dx = 0.3699$$

In this case,

$$\beta = 1 - \sum_{x=0}^{2} \binom{12}{x} (0.3699)^x (0.6301)^{12-x} = 0.8794$$

Thus, β for the sign test depends not only on the alternative value of $\tilde{\mu}$ but also on the area to the right of the value specified in the null hypothesis under the population probability distribution. This area depends highly on the shape of that particular probability distribution. In this example, β is large, so the ability of the test to detect this departure from the null hypothesis with the current sample size is poor.

9.9.2 The Wilcoxon Signed-Rank Test

The sign test uses only the plus and minus signs of the differences between the observations and the median $\tilde{\mu}_0$ (or the plus and minus signs of the differences between the observations in the paired case). It does not take into account the size or magnitude of these differences. Frank Wilcoxon devised a test procedure that uses both direction (sign) and magnitude. This procedure, now called the **Wilcoxon signed-rank test**, is discussed and illustrated in this section.

The Wilcoxon signed-rank test applies to the case of **symmetric continuous distributions**. Under these assumptions, the mean equals the median, and we can use this procedure to test the null hypothesis $\mu = \mu_0$.

The Test Procedure We are interested in testing H_0: $\mu = \mu_0$ against the usual alternatives. Assume that $X_1, X_2, \ldots, X_n$ is a random sample from a continuous and symmetric distribution with mean (and median) μ. Compute the differences $X_i - \mu_0$, $i = 1, 2, \ldots, n$. Rank the absolute differences $|X_i - \mu_0|$, $i = 1, 2, \ldots, n$ in ascending order, and then give the ranks the signs of their corresponding differences. Let W^+ be the sum of the positive ranks and W^- be the absolute value of the sum of the negative ranks, and let $W = \min(W^+, W^-)$. Appendix Table IX contains critical values of W, say, W^*_α. If the alternative hypothesis is H_1: $\mu \neq \mu_0$, then if the observed value of the statistic $w \leq w^*_\alpha$, the null hypothesis H_0: $\mu = \mu_0$ is rejected. Appendix Table IX provides significance levels of $\alpha = 0.10$, $\alpha = 0.05$, $\alpha = 0.02$, and $\alpha = 0.01$ for the two-sided test.

For one-sided tests, if the alternative is H_1: $\mu > \mu_0$, reject H_0: $\mu = \mu_0$ if $w^- \leq w^*_\alpha$; and if the alternative is H_1: $\mu < \mu_0$, reject H_0: $\mu = \mu_0$ if $w^+ \leq w^*_\alpha$. The significance levels for one-sided tests provided in Appendix Table IX are $\alpha = 0.05$, 0.025, 0.01, and 0.005.

EXAMPLE 9.16 | Propellant Shear Strength-Wilcoxon Signed-Rank Test

We illustrate the Wilcoxon signed-rank test by applying it to the propellant shear strength data from Table 9.5. Assume that the underlying distribution is a continuous symmetric distribution. The seven-step procedure is applied as follows:

1. **Parameter of interest:** The parameter of interest is the mean (or median) of the distribution of propellant shear strength.
2. **Null hypothesis:** H_0: $\mu = 2000$ psi
3. **Alternative hypothesis:** H_0: $\mu \neq 2000$ psi
4. **Test statistic:** The test statistic is $w = \min(w^+, w^-)$
5. **Reject H_0 if:** We will reject H_0 if $w \leq w^*_{0.05} = 52$ from Appendix Table IX.
6. **Computations:** The signed ranks from Table 9.5 are shown in the following display:

   The sum of the positive ranks is $w^+ = (1 + 2 + 3 + 4 + 5 + 6 + 11 + 13 + 15 + 16 + 17 + 18 + 19 + 20) = 150$, and the sum of the absolute values of the negative ranks is $w^- = (7 + 8 + 9 + 10 + 12 + 14) = 60$. Therefore,

   $$w = \min(150, 60) = 60$$

Observation	Difference $x_i - 2000$	Signed Rank
16	+53.50	+1
4	+61.30	+2
1	+158.70	+3
11	+165.20	+4
18	+200.50	+5
5	+207.50	+6
7	+215.30	−7
13	−220.20	−8
15	−234.70	−9
20	−246.30	−10
10	+256.70	+11
6	−291.70	−12
3	+316.00	+13
2	−321.85	−14
14	+336.75	+15
9	+357.90	+16
12	+399.55	+17
17	+414.40	+18
8	+575.10	+19
19	+654.20	+20

7. **Conclusions:** Because $w = 60$ is not less than or equal to the critical value $w_{0.05} = 52$, we cannot reject the null hypothesis that the mean (or median, because the population is assumed to be symmetric) shear strength is 2000 psi.

Ties in the Wilcoxon Signed-Rank Test Because the underlying population is continuous, ties are theoretically impossible, although they will sometimes occur in practice. If several observations have the same absolute magnitude, they are assigned the average of the ranks that they would receive if they differed slightly from one another.

Large Sample Approximation If the sample size is moderately large, say, $n > 20$, it can be shown that W^+ (or W^-) has approximately a normal distribution with mean

$$\mu_{W^+} = \frac{n(n+1)}{4}$$

and variance

$$\sigma^2_{W^+} = \frac{n(n+1)(2n+1)}{24}$$

Therefore, a test of H_0: $\mu = \mu_0$ can be based on the statistic:

Normal Approximation for Wilcoxon Signed-Rank Statistic

$$Z_0 = \frac{W^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \tag{9.56}$$

An appropriate critical region for either the two-sided or one-sided alternative hypotheses can be chosen from a table of the standard normal distribution.

9.9.3 Comparison to the *t*-Test

If the underlying population is normal, either the sign test or the *t*-test could be used to test a hypothesis about the population median. The *t*-test is known to have the smallest value of β possible among all tests that have significance level α for the one-sided alternative and for tests with symmetric critical regions for the two-sided alternative, so it is superior to the sign test in the normal distribution case. When the population distribution is symmetric and nonnormal (but with finite mean), the *t*-test will have a smaller β (or a higher power) than the sign test unless the distribution has very heavy tails compared with the normal. Thus, the sign test is usually considered a test procedure for the median rather than as a serious competitor for the *t*-test. The Wilcoxon signed-rank test is preferable to the sign test and compares well with the *t*-test for symmetric distributions. It can be useful for situations in which a transformation on the observations does not produce a distribution that is reasonably close to the normal.

9.10 Equivalence Testing

Statistical hypothesis testing is one of the most useful techniques of statistical inference. However, it works in only one direction; that is, it starts with a statement that is assumed to be true (the null hypothesis H_0) and attempts to disprove this claim in favor of the alternative hypothesis H_1. The strong statement about the alternative hypothesis is made when the null hypothesis is rejected. This procedure works well in many but not all situations.

To illustrate, consider a situation in which we are trying to qualify a new supplier of a component that we use in manufacturing our product. The current supplier produces these components with a standard mean resistance of 80 ohms. If the new supplier can provide components with the same mean resistance, we will qualify them. Having a second source for this component is considered to be important because demand for our product is expected to grow rapidly in the

near future, and the second supplier will be necessary to meet the anticipated increase in demand. The traditional formulation of the hypothesis test

$$H_0: \mu = 80 \qquad H_1: \mu \neq 80$$

really is not satisfactory. Only if we reject the null hypothesis do we have a strong conclusion. We actually want to state the hypotheses as follows:

$$H_0: \mu \neq 80 \qquad H_1: \mu = 80$$

This type of hypothesis statement is called an **equivalence test**. We assume that the new supplier is different from the standard unless we have strong evidence to reject that claim. The way that this equivalence test is carried out is to test the following two sets of one-sided alternative hypotheses:

$$H_0: \mu = 80 + \delta \qquad H_1: \mu < 80 + \delta$$

and

$$H_0: \mu = 80 - \delta \qquad H_1: \mu > 80 - \delta$$

where δ is called the **equivalence band**, which is a practical threshold or limit within which the mean performance (here the resistance) is considered to be the same as the standard. The interval $80 \pm \delta$ is called an **equivalence interval**. The first set of hypotheses is a test of the mean that shows that the difference between the mean and the standard is significantly less than the upper equivalence limit of the interval, and the second set of hypotheses is a test of the mean that shows that the difference between the mean and the standard is significantly greater than the lower equivalence limit. We are going to apply both tests to the same sample of data, leading to a test of equivalence that is sometimes called **two one-sided tests (TOST)**.

EXAMPLE 9.17

Suppose that we have a random sample of $n = 50$ components from the new supplier. Resistance is approximately normally distributed, and the sample mean and standard deviation (in ohms) are $\bar{x} = 79.98$ and $s = 0.10$. The sample mean is close to the standard of 80 ohms. Suppose that our error of measurement is approximately 0.01 ohm. We will decide that if the new supplier has a mean resistance that is within 0.05 of the standard of 80, there is no practical difference in performance. Therefore, $\delta = 0.05$. Notice that we have chosen the equivalence band to be greater than the usual or expected measurement error for the resistance. We now want to test the hypotheses

$$H_0: \mu = 80.05 \qquad H_1: \mu < 80.05$$

and

$$H_0: \mu = 79.95 \qquad H_1: \mu > 79.95$$

Consider testing the first set of hypotheses. It is straightforward to show that the value of the test statistic is $t_0 = -4.95$, and the P-value is less than 0.01. Therefore, we conclude that the mean resistance is less than 80.05. For the second set of hypotheses, the test statistic is $t_0 = 2.12$, and the P-value is less than 0.025, so the mean resistance is significantly greater than 79.95 and significantly less than 80.05. Thus, we have enough evidence to conclude that the new supplier produces components that are equivalent to those produced by the current supplier because the mean is within the ± 0.05 ohm interval.

Equivalence testing has many applications, including the supplier qualification problem illustrated here, generic drug manufacturing, and new equipment qualification. The experimenter must decide what defines equivalence. Issues that should be considered include these:

1. Specifying the equivalence band. The parameter δ should be larger than the typical measurement error. A good rule of thumb is that δ should be at least three times the typical measurement error.

2. The equivalence band should be much smaller than the usual process variation.
3. The equivalence band should be much smaller than the product or process specifications. Specifications typically define fitness for use.
4. The equivalence band should be related to actual functional performance; that is, how much of a difference can be tolerated before performance is degraded?

9.11 Combining *P*-Values

Testing several sets of hypotheses that relate to a problem of interest occurs fairly often in engineering and many scientific disciplines. For example, suppose that we are developing a new synthetic fiber to be used in manufacturing body armor for the military and law enforcement agencies. This fiber needs to exhibit a high breaking strength (at least 100 lb/in^2) for the new product to work properly. The engineering development lab produced several batches or lots of this fiber, a random sample of three fiber specimens from each lot has been taken, and the sample specimens tested. For each lot, the hypotheses of interest are

$$H_0: \mu = 100 \qquad H_1: \mu > 100$$

The development lots are small, and the testing is destructive, so the sample sizes are also small. After six lots have been produced, the *P*-values from these six independent tests of hypotheses are 0.105, 0.080, 0.250, 0.026, 0.650, and 0.045. Given the size of these *P*-values, we suspect that the new material is going to be satisfactory, but the sample sizes are small, and it would be helpful if we could combine the results from all six tests to determine whether the new material will be acceptable. Combining results from several studies or experiments is sometimes called **meta-analysis**, a technique that has been used in many fields including public health monitoring, clinical trials of new medical devices or treatments, ecology, and genetics. One method that can be used to combine these results is to combine all of the individual *P*-values into a single statistic for which one *P*-value can be computed. This procedure was developed by R. A. Fisher.

Let P_i be the *P*-value for the *i*th set of hypotheses, $i = 1, 2, \ldots, m$. The test statistic is

$$\chi_0^2 = -2\sum_{i=1}^{m} \ln(P_i)$$

The test statistic χ_0^2 follows a chi-square distribution with $2m$ degrees of freedom. A *P*-value can be computed for the observed value of this statistic. A small *P*-value would lead to rejection of the shared null hypotheses and a conclusion that the combined data support the alternative.

As an example, the test statistic χ_0^2 for the six tests described is

$$\chi_0^2 = -2[\ln(0.105) + \ln(0.080) + \ln(0.250) + \ln(0.026) + \ln(0.650) + \ln(0.045)] = 26.6947$$

with $2m = 2(6) = 12$ degrees of freedom. The *P*-value for this statistic is $0.005 < P < 0.01$, a very small value, which leads to rejection of the null hypothesis. In other words, the combined information from all six tests provides evidence that the mean fiber strength exceeds 100 lb/in^2.

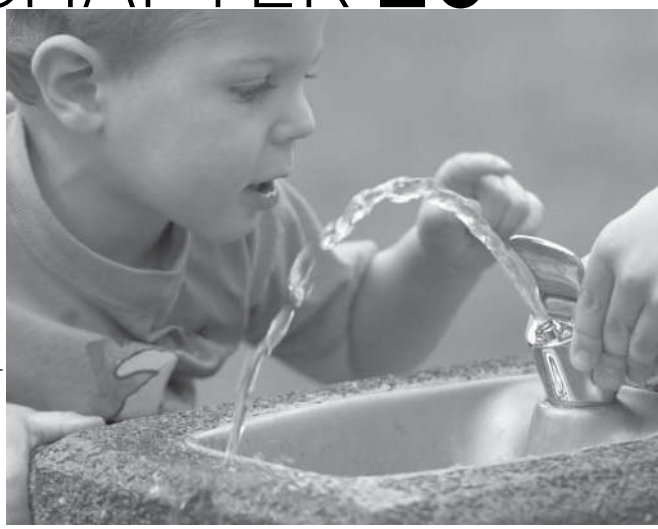
Fisher's method does not require all the null hypotheses be the same. Some applications involve many sets of hypotheses that do not have the same null. In these situations, the alternative hypothesis is taken to be that at least one of the null hypotheses is false. Fisher's method was developed in the 1920s. Since then, a number of other techniques has been proposed. For a good discussion of these alternative methods along with comments on their appropriateness and power, see the article by Piegorsch and Bailer ["Combining Information," *Wiley Interdisciplinary Reviews: Computational Statistics,* 2009, Vol. 1(3), pp. 354–360].

Important Terms and Concepts

Acceptance region
Alternative hypothesis
α and β
Chi-square tests
Combining *P*-values
Confidence interval
Connection between hypothesis tests and confidence intervals
Contingency table
Critical region for a test statistic
Critical values
Equivalence testing
Fixed significance level
Goodness-of-fit test
Homogeneity test
Hypotheses
Hypothesis testing
Independence test
Inference
Nonparametric and distribution free methods
Normal approximation to nonparametric tests
Null distribution
Null hypothesis
Observed significance level
One- and two-sided alternative hypotheses
Operating characteristic (OC) curves
Parametric
Power of a statistical test
P-value
Ranks
Reference distribution for a test statistic
Rejection region
Sampling distribution
Sample size determination for hypothesis tests
Sign test
Significance level of a test
Statistical hypothesis
Statistical versus practical significance
Symmetric continuous distributions
t-test
Test statistic
Type I and type II errors
Wilcoxon's signed-rank test
z-test

CHAPTER **10**

Statistical Inference for Two Samples



© HKPNC / iStockphoto

LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Structure comparative experiments involving two samples as hypothesis tests
2. Test hypotheses and construct confidence intervals on the difference in means of two normal distributions
3. Test hypotheses and construct confidence intervals on the ratio of the variances or standard deviations of two normal distributions
4. Test hypotheses and construct confidence intervals on the difference in two population proportions
5. Use the *P*-value approach for making decisions in hypotheses tests
6. Compute power, and type II error probability, and make sample size decisions for two-sample tests on means, variances, and proportions
7. Explain and use the relationship between confidence intervals and hypothesis tests

CHAPTER OUTLINE

10.1 Inference on the Difference in Means of Two Normal Distributions, Variances Known

- 10.1.1 Hypothesis Tests on the Difference in Means, Variances Known
- 10.1.2 Type II Error and Choice of Sample Size
- 10.1.3 Confidence Interval on the Difference in Means, Variances Known

10.2 Inference on the Difference in Means of Two Normal Distributions, Variances Unknown

- 10.2.1 Hypotheses Tests on the Difference in Means, Variances Unknown
- 10.2.2 Type II Error and Choice of Sample Size
- 10.2.3 Confidence Interval on the Difference in Means, Variances Unknown

10.3 A Nonparametric Test for the Difference in Two Means

- 10.3.1 Description of the Wilcoxon Rank-Sum Test
- 10.3.2 Large-Sample Approximation
- 10.3.3 Comparison to the *t*-Test

10.4 Paired *t*-Test

10.5 Inference on the Variances of Two Normal Distributions

- 10.5.1 *F* Distribution
- 10.5.2 Hypothesis Tests on the Equity of Two Variances
- 10.5.3 Type II Error and Choice of Sample Size
- 10.5.4 Confidence Interval on the Ratio of Two Variances

10.6 Inference on Two Population Proportions

- 10.6.1 Large-Sample Tests on the Difference in Population Proportions
- 10.6.2 Type II Error and Choice of Sample Size
- 10.6.3 Confidence Interval on the Difference in Population Proportions

10.7 Summary Table and Road Map for Inference Procedures for Two Samples

The safety of drinking water is a serious public health issue. An article in the *Arizona Republic* on May 27, 2001, reported on arsenic contamination in the water sampled from 10 communities in the metropolitan Phoenix area and 10 communities from rural Arizona. The data showed dramatic differences in the arsenic concentration, ranging from 3 parts per billion (ppb) to 48 ppb. This article suggested some important questions. Does real difference in the arsenic concentrations in the Phoenix area and in the rural communities in Arizona exist? How large is this difference? Is it large enough to require action on the part of the public health service and other state agencies to correct the problem? Are the levels of reported arsenic concentration large enough to constitute a public health risk?

Some of these questions can be answered by statistical methods. If we think of the metropolitan Phoenix communities as one population and the rural Arizona communities as a second population, we could determine whether a statistically significant difference in the mean arsenic concentration exists for the two populations by testing the hypothesis that the two means, say, μ_1 and μ_2, are different. This is a relatively simple extension to two samples of the one-sample hypothesis testing procedures of Chapter 9. We could also use a confidence interval to estimate the difference in the two means, say, $\mu_1 - \mu_2$.

The arsenic concentration problem is very typical of many problems in engineering and science that involve statistics. Some of the questions can be answered by the application of appropriate statistical tools, and other questions require using engineering or scientific knowledge and expertise to answer satisfactorily.

10.1 Inference on the Difference in Means of Two Normal Distributions, Variances Known

The previous two chapters presented hypothesis tests and confidence intervals for a single population parameter (the mean μ, the variance σ^2, or a proportion p). This chapter extends those results to the case of two independent populations.

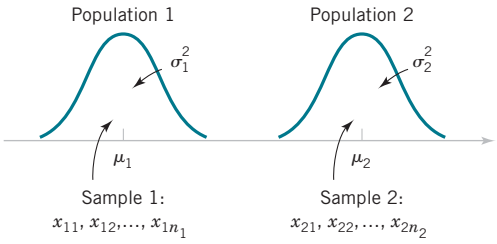


FIGURE 10.1
Two independent populations.

The general situation is shown in Figure 10.1. Population 1 has mean μ_1 and variance σ_1^2, and population 2 has mean μ_2 and variance σ_2^2. Inferences will be based on two random samples of sizes n_1 and n_2, respectively. That is, $X_{11}, X_{12}, \ldots, X_{1n_1}$ is a random sample of n_1 observations from population 1, and $X_{21}, X_{22}, \ldots, X_{2n_2}$ is a random sample of n_2 observations from population 2. Most of the practical applications of the procedures in this chapter arise in the context of simple **comparative experiments** in which the objective is to study the difference in the parameters of the two populations.

Engineers and scientists are often interested in comparing two different conditions to determine whether either condition produces a significant effect on the response that is observed. These conditions are sometimes called **treatments**. Example 10.1 describes such an experiment; the two different treatments are two paint formulations, and the response is the drying time. The purpose of the study is to determine whether the new formulation results in a significant effect—reducing drying time. In this situation, the product developer (the experimenter) randomly assigned 10 test specimens to one formulation and 10 test specimens to the other formulation. Then the paints were applied to the test specimens in random order until all 20 specimens were painted. This is an example of a **completely randomized experiment**.

When statistical significance is observed in a randomized experiment, the experimenter can be confident in the conclusion that the difference in treatments resulted in the difference in response. That is, we can be confident that a **cause-and-effect** relationship has been found.

Sometimes the objects to be used in the comparison are not assigned at random to the treatments. For example, the September 1992 issue of *Circulation* (a medical journal published by the American Heart Association) reports a study linking high iron levels in the body with increased risk of heart attack. The study, done in Finland, tracked 1931 men for 5 years and showed a statistically significant effect of increasing iron levels on the incidence of heart attacks. In this study, the comparison was not performed by randomly selecting a sample of men and then assigning some to a "low iron level" treatment and the others to a "high iron level" treatment. The researchers just tracked the subjects over time. Recall from Chapter 1 that this type of study is called an **observational study**.

It is difficult to identify causality in observational studies because the observed statistically significant difference in response for the two groups may be due to some other underlying factor (or group of factors) that was not equalized by randomization and not due to the treatments. For example, the difference in heart attack risk could be attributable to the difference in iron levels or to other underlying factors that form a reasonable explanation for the observed results—such as cholesterol levels or hypertension.

In this section, we consider statistical inferences on the difference in means $\mu_1 - \mu_2$ of two normal distributions where the variances σ_1^2 and σ_2^2 are known. The assumptions for this section are summarized as follows.

Assumptions for Two-Sample Inference

(1) $X_{11}, X_{12}, \ldots, X_{1n_1}$ is a random sample from population 1.
(2) $X_{21}, X_{22}, \ldots, X_{2n_2}$ is a random sample from population 2.

(3) The two populations represented by X_1 and X_2 are independent.
(4) Both populations are normal.

A logical point estimator of $\mu_1 - \mu_2$ is the difference in sample means $\overline{X}_1 - \overline{X}_2$. Based on the properties of expected values,

$$E(\overline{X}_1 - \overline{X}_2) = E(\overline{X}_1) - E(\overline{X}_2) = \mu_1 - \mu_2$$

and the variance of $\overline{X}_1 - \overline{X}_2$ is

$$V(\overline{X}_1 - \overline{X}_2) = V(\overline{X}_1) + V(\overline{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Based on the assumptions and the preceding results, we may state the following.

The quantity

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{10.1}$$

has a $N(0, 1)$ distribution.

This result is used to develop procedures for tests of hypotheses and to construct confidence intervals on $\mu_1 - \mu_2$. Essentially, we may think of $\mu_1 - \mu_2$ as a parameter θ where estimator is $\hat{\Theta} = \overline{X}_1 - \overline{X}_2$ with variance $\sigma^2_{\hat{\Theta}} = \sigma_1^2/n_1 + \sigma_2^2/n_2$. If θ_0 is the null hypothesis value specified for θ, the test statistic will be $(\hat{\Theta} - \theta_0)/\sigma_{\hat{\Theta}}$. Notice how similar this is to the test statistic for a single mean used in Equation 9.8 in Chapter 9.

10.1.1 Hypothesis Tests on the Difference in Means, Variances Known

We now consider hypothesis testing on the difference in the means $\mu_1 - \mu_2$ of two normal populations. Suppose that we are interested in testing whether the difference in means $\mu_1 - \mu_2$ is equal to a specified value Δ_0. Thus, the null hypothesis is stated as $H_0: \mu_1 - \mu_2 = \Delta_0$. Obviously, in many cases, we specify $\Delta_0 = 0$ so that we are testing the equality of two means (i.e., $H_0: \mu_1 = \mu_2$). The appropriate test statistic would be found by replacing $\mu_1 - \mu_2$ in Equation 10.1 by Δ_0: this test statistic would have a standard normal distribution under H_0. That is, the standard normal distribution is the *reference distribution* for the test statistic. Suppose that the alternative hypothesis is $H_1: \mu_1 - \mu_2 \neq \Delta_0$. A sample value of $\overline{x}_1 - \overline{x}_2$ that is considerably different from Δ_0 is evidence that H_1 is true. Because Z_0 has the $N(0, 1)$ distribution when H_0 is true, we would calculate the P-value as the sum of the probabilities beyond the test statistic value z_0 and $-z_0$ in the standard normal distribution. That is, $P = 2[1 - \Phi(|z_0|)]$. This is exactly what we did in the one-sample z-test. If we wanted to perform a fixed-significance-level test, we would take $-z_{\alpha/2}$ and $z_{\alpha/2}$ as the boundaries of the critical region just as we did in the single-sample z-test. This would give a test with level of significance α. P-values or critical regions for the one-sided alternatives would be determined similarly. Formally, we summarize these results in the following display.

Tests on the Difference in Means, Variances Known

Null hypothesis: $H_0: \mu_1 - \mu_2 = \Delta_0$

Test statistic:
$$Z_0 = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{10.2}$$

Alternative Hypotheses	*P*-Value	Rejection Criterion for Fixed-Level Tests
$H_1: \mu_1 - \mu_2 \neq \Delta_0$	Probability above $\lvert z_0 \rvert$ and probability below $-\lvert z_0 \rvert$, $P = 2[1 - \Phi(\lvert z_0 \rvert)]$	$z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$
$H_1: \mu_1 - \mu_2 > \Delta_0$	Probability above z_0, $P = 1 - \Phi(z_0)$	$z_0 > z_\alpha$
$H_1: \mu_1 - \mu_2 < \Delta_0$	Probability below z_0, $P = \Phi(z_0)$	$z_0 < -z_\alpha$

EXAMPLE 10.1 | Paint Drying Time

A product developer is interested in reducing the drying time of a primer paint. Two formulations of the paint are tested; formulation 1 is the standard chemistry, and formulation 2 has a new drying ingredient that should reduce the drying time. From experience, it is known that the standard deviation of drying time is 8 minutes, and this inherent variability should be unaffected by the addition of the new ingredient. Ten specimens are painted with formulation 1, and another 10 specimens are painted with formulation 2; the 20 specimens are painted in random order. The two sample average drying times are $\bar{x}_1 = 121$ minutes and $\bar{x}_2 = 112$ minutes, respectively. What conclusions can the product developer draw about the effectiveness of the new ingredient, using $\alpha = 0.05$?

We apply the seven-step procedure to this problem as follows:

1. **Parameter of interest:** The quantity of interest is the difference in mean drying times, $\mu_1 - \mu_2$, and $\Delta_0 = 0$.
2. **Null hypothesis:** $H_0: \mu_1 - \mu_2 = 0$, or $H_0: \mu_1 = \mu_2$.
3. **Alternative hypothesis:** $H_1: \mu_1 > \mu_2$. We want to reject H_0 if the new ingredient reduces mean drying time.
4. **Test statistic:** The test statistic is
$$z_0 = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$
where $\sigma_1^2 = \sigma_2^2 = (8)^2 = 64$ and $n_1 = n_2 = 10$.
5. **Reject H_0 if:** Reject $H_0: \mu_1 = \mu_2$ if the *P*-value is less than 0.05.
6. **Computations:** Because $\bar{x}_1 = 121$ minutes and $\bar{x}_2 = 112$ minutes, the test statistic is
$$z_0 = \frac{121 - 112}{\sqrt{\frac{(8)^2}{10} + \frac{(8)^2}{10}}} = 2.52$$
7. **Conclusion:** Because $z_0 = 2.52$, the *P*-value is $P = 1 - \Phi(2.52) = 0.0059$, so we reject H_0 at the $\alpha = 0.05$ level.

Practical Interpretation: We conclude that adding the new ingredient to the paint significantly reduces the drying time. This is a strong conclusion.

When the population variances are unknown, the sample variances s_1^2 and s_2^2 can be substituted into the test statistic Equation 10.2 to produce a *large-sample test* for the difference in means. This procedure will also work well when the populations are not necessarily normally distributed. However, both n_1 and n_2 should exceed 40 for this large-sample test to be valid.

10.1.2 Type II Error and Choice of Sample Size

Use of Operating Characteristic Curves The operating characteristic (OC) curves in Appendix Charts VII*a*, VII*b*, VII*c*, and VII*d* may be used to evaluate the type II error probability for the hypotheses in the display (10.2). These curves are also useful in determining sample size. Curves are provided for $\alpha = 0.05$ and $\alpha = 0.01$. For the two-sided alternative hypothesis, the abscissa scale of the operating characteristic curve in charts VII*a* and VII*b* is d, where

$$d = \frac{|\mu_1 - \mu_2 - \Delta_0|}{\sqrt{\sigma_1^2 + \sigma_2^2}} = \frac{|\Delta - \Delta_0|}{\sqrt{\sigma_1^2 + \sigma_2^2}} \tag{10.3}$$

and one must choose equal sample sizes, say, $n = n_1 = n_2$. The one-sided alternative hypotheses require the use of Charts VII*c* and VII*d*. For the one-sided alternatives $H_1: \mu_1 - \mu_2 > \Delta_0$ or $H_1: \mu_1 - \mu_2 < \Delta_0$, the abscissa scale is also given by

$$d = \frac{|\mu_1 - \mu_2 - \Delta_0|}{\sqrt{\sigma_1^2 + \sigma_2^2}} = \frac{|\Delta - \Delta_0|}{\sqrt{\sigma_1^2 + \sigma_2^2}}$$

It is not unusual to encounter problems where the costs of collecting data differ substantially for the two populations or when the variance for one population is much greater than the other. In those cases, we often use unequal sample sizes. If $n_1 \neq n_2$, the operating characteristic curves may be entered with an *equivalent* value of n computed from

$$n = \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} \tag{10.4}$$

If $n_1 \neq n_2$ and their values are fixed in advance, Equation 10.4 is used directly to calculate n, and the operating characteristic curves are entered with a specified d to obtain β. If we are given d and it is necessary to determine n_1 and n_2 to obtain a specified β, say, β^*, we guess at trial values of n_1 and n_2, calculate n in Equation 10.4, and enter the curves with the specified value of d to find β. If $\beta = \beta^*$, the trial values of n_1 and n_2 are satisfactory. If $\beta \neq \beta^*$, adjustments to n_1 and n_2 are made and the process is repeated.

EXAMPLE 10.2 | Paint Drying Time, Sample Size from OC Curves

Consider the paint drying time experiment from Example 10.1. If the true difference in mean drying times is as much as 10 minutes, find the sample sizes required to detect this difference with probability at least 0.90.

The appropriate value of the abscissa parameter is (because $\Delta_0 = 0$, and $\Delta = 10$)

$$d = \frac{|\mu_1 - \mu_2|}{\sqrt{\sigma_1^2 + \sigma_2^2}} = \frac{10}{\sqrt{8^2 + 8^2}} = 0.88$$

and because the detection probability or power of the test must be at least 0.9, with $\alpha = 0.05$, we find from Appendix Chart VII*c* that $n = n_1 = n_2 \simeq 11$.

Sample Size Formulas It is also possible to obtain formulas for calculating the sample sizes directly. Suppose that the null hypothesis $H_0: \mu_1 - \mu_2 = \Delta_0$ is false and that the true difference in means is $\mu_1 - \mu_2 = \Delta$ where $\Delta > \Delta_0$. One may find formulas for the sample size required to

obtain a specific value of the type II error probability β for a given difference in means Δ and level of significance α.

For example, we first write the expression for the β-error for the two-sided alternative, which is

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right) - \Phi\left(-z_{\alpha/2} - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right)$$

The derivation for sample size closely follows the single-sample case in Section 9.2.2.

Sample Size for a Two-Sided Test on the Difference in Means with $n_1 = n_2$, Variances Known

For the two-sided alternative hypothesis with significance level α, the sample size $n_1 = n_2 = n$ required to detect a true difference in means of Δ with power at least $1 - \beta$ is

$$n \simeq \frac{\left(z_{\alpha/2} + z_\beta\right)^2(\sigma_1^2 + \sigma_2^2)}{(\Delta - \Delta_0)^2} \tag{10.5}$$

This approximation is valid when $\Phi\left(-z_{\alpha/2} - (\Delta - \Delta_0)\sqrt{n}/\sqrt{\sigma_1^2 + \sigma_2^2}\right)$ is small compared to β.

Sample Size for a One-Sided Test on the Difference in Means with $n_1 = n_2$, Variances Known

For a one-sided alternative hypothesis with significance level α, the sample size $n_1 = n_2 = n$ required to detect a true difference in means of $\Delta(\neq \Delta_0)$ with power at least $1 - \beta$ is

$$n = \frac{(z_\alpha + z_\beta)^2\,(\sigma_1^2 + \sigma_2^2)}{(\Delta - \Delta_0)^2} \tag{10.6}$$

where Δ is the true difference in means of interest. Then by following a procedure similar to that used to obtain Equation 9.17, the expression for β can be obtained for the case where $n = n_1 = n_2$.

EXAMPLE 10.3 | Paint Drying Time Sample Size

To illustrate the use of these sample size equations, consider the situation described in Example 10.1, and suppose that if the true difference in drying times is as much as 10 minutes, we want to detect this with probability at least 0.90. Under the null hypothesis, $\Delta_0 = 0$. We have a one-sided alternative hypothesis with $\Delta = 10$, $\alpha = 0.05$ (so $z_\alpha = z_{0.05} = 1.645$), and because the power is 0.9, $\beta = 0.10$ (so $z_\beta = z_{0.10} = 1.28$). Therefore, we may find the required sample size from Equation 10.6 as follows:

$$n = \frac{(z_\alpha + z_\beta)^2\,(\sigma_1^2 + \sigma_2^2)}{(\Delta - \Delta_0)^2} = \frac{(1.645 + 1.28)^2\,[(8)^2 + (8)^2]}{(10 - 0)^2} \approx 11$$

This is exactly the same as the result obtained from using the OC curves.

10.1.3 Confidence Interval on the Difference in Means, Variances Known

The $100(1-\alpha)\%$ confidence interval on the difference in two means $\mu_1 - \mu_2$ when the variances are known can be found directly from results given previously in this section. Recall that $X_{11}, X_{12}, \ldots, X_{1n_1}$ is a random sample of n_1 observations from the first population and $X_{21}, X_{22}, \ldots, X_{2n_2}$ is a random sample of n_2 observations from the second population. The difference in sample means $\overline{X}_1 - \overline{X}_2$ is a point estimator of $\mu_1 - \mu_2$, and

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

has a standard normal distribution if the two populations are normal or is approximately standard normal if the conditions of the central limit theorem apply, respectively. This implies that $P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1 - \alpha$, or

$$P\left[-z_{\alpha/2} \le \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \le z_{\alpha/2}\right] = 1 - \alpha$$

This can be rearranged as

$$P\left(\overline{X}_1 - \overline{X}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le \overline{X}_1 - \overline{X}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha$$

Therefore, the $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is defined as follows.

Confidence Interval on the Difference in Means, Variances Known

If $\bar{x}_1$ and $\bar{x}_2$ are the means of independent random samples of sizes n_1 and n_2 from two independent normal populations with known variances σ_1^2 and σ_2^2, respectively, a $100(1-\alpha)\%$ **confidence interval for $\mu_1 - \mu_2$** is

$$\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad (10.7)$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution.

The confidence level $1-\alpha$ is exact when the populations are normal. For nonnormal populations, the confidence level is approximately valid for large sample sizes.

Equation 10.7 can also be used as a *large sample CI* on the difference in mean when σ_1^2 and σ_2^2 are unknown by substituting s_1^2 and s_2^2 for the population variances. For this to be a valid procedure, both sample sizes n_1 and n_2 should exceed 40.

EXAMPLE 10.4 | Aluminum Tensile Strength

Tensile strength tests were performed on two different grades of aluminum spars used in manufacturing the wing of a commercial transport aircraft. From past experience with the spar manufacturing process and the testing procedure, the standard deviations of tensile strengths are assumed to be known. The data obtained are as follows: $n_1 = 10$, $\bar{x}_1 = 87.6, \sigma_1 = 1, n_2 = 12$, $\bar{x}_2 = 74.5$, and $\sigma_2 = 1.5$. If μ_1 and μ_2 denote the true mean tensile strengths for the two grades of spars, we may find a 90% CI on the difference in mean strength $\mu_1 - \mu_2$ as follows:

$$\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$87.6 - 74.5 - 1.645\sqrt{\frac{(1)^2}{10} + \frac{(1.5)^2}{12}} \le \mu_1 - \mu_2 \le 87.6 - 74.5 + 1.645\sqrt{\frac{(1^2)}{10} + \frac{(1.5)^2}{12}}$$

Therefore, the 90% confidence interval on the difference in mean tensile strength (in kilograms per square millimeter) is

$$12.22 \le \mu_1 - \mu_2 \le 13.98 \text{ (in kilograms per square millimeter)}$$

Practical Interpretation: Notice that the confidence interval does not include zero, implying that the mean strength of aluminum grade 1 (μ_1) exceeds the mean strength of aluminum grade 2 (μ_2). In fact, we can state that we are 90% confident that the mean tensile strength of aluminum grade 1 exceeds that of aluminum grade 2 by between 12.22 and 13.98 kilograms per square millimeter.

Choice of Sample Size If the standard deviations σ_1 and σ_2 are known (at least approximately) and the two sample sizes n_1 and n_2 are equal ($n_1 = n_2 = n$, say), we can determine the sample size required so that the error in estimating $\mu_1 - \mu_2$ by $\bar{x}_1 - \bar{x}_2$ will be less than E at $100(1-\alpha)\%$ confidence. The required sample size from each population is

Sample Size for a Confidence Interval on the Difference in Means, Variances Known

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 (\sigma_1^2 + \sigma_2^2) \tag{10.8}$$

Remember to round up if n is not an integer. This ensures that the level of confidence does not drop below $100(1-\alpha)\%$.

One-Sided Confidence Bounds One-sided confidence bounds on $\mu_1 - \mu_2$ may also be obtained. A $100(1-\alpha)\%$ upper-confidence bound on $\mu_1 - \mu_2$ is

One-Sided Upper-Confidence Bound

$$\mu_1 - \mu_2 \le \bar{x}_1 - \bar{x}_2 + z_\alpha\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{10.9}$$

and a $100(1-\alpha)\%$ lower-confidence bound is

One-Sided Lower-Confidence Bound

$$\bar{x}_1 - \bar{x}_2 - z_\alpha\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \tag{10.10}$$

10.2 Inference on the Difference in Means of Two Normal Distributions, Variances Unknown

We now extend the results of the previous section to the difference in means of the two distributions in Figure 10.1 when the variances of both distributions σ_1^2 and σ_2^2 are unknown. If the sample sizes n_1 and n_2 exceed 40, the normal distribution procedures in Section 10.1 could be used. However, when small samples are taken, we assume that the populations are normally distributed and base our hypotheses tests and confidence intervals on the t distribution. This parallels nicely the case of inference on the mean of a single sample with unknown variance.

10.2.1 Hypotheses Tests on the Difference in Means, Variances Unknown

We now consider tests of hypotheses on the difference in means $\mu_1 - \mu_2$ of two normal distributions where the variances σ_1^2 and σ_2^2 are unknown. A t-statistic is used to test these hypotheses. As noted earlier and in Section 9.3, the normality assumption is required to develop the test procedure, but moderate departures from normality do not adversely affect the procedure. Two different situations must be treated. In the first case, we assume that the variances of the two normal distributions are unknown but equal; that is, $\sigma_1^2 = \sigma_2^2 = \sigma^2$. In the second, we assume that σ_1^2 and σ_2^2 are unknown and not necessarily equal.

Case 1: $\sigma_1^2 = \sigma_2^2 = \sigma^2$ Suppose that we have two independent normal populations with unknown means μ_1 and μ_2, and unknown but equal variances, $\sigma_1^2 = \sigma_2^2 = \sigma^2$. We wish to test

$$\begin{aligned} H_0&: \mu_1 - \mu_2 = \Delta_0 \\ H_1&: \mu_1 - \mu_2 \neq \Delta_0 \end{aligned} \tag{10.11}$$

Let $X_{11}, X_{12}, \ldots, X_{1n_1}$ be a random sample of n_1 observations from the first population and $X_{21}, X_{22}, \ldots, X_{2n_2}$ be a random sample of n_2 observations from the second population. Let $\overline{X}_1, \overline{X}_2, S_1^2$, and S_2^2 be the sample means and sample variances, respectively. Now the expected value of the difference in sample means $\overline{X}_1 - \overline{X}_2$ is $E(\overline{X}_1 - \overline{X}_2) = \mu_1 - \mu_2$, so $\overline{X}_1 - \overline{X}_2$ is an unbiased estimator of the difference in means. The variance of $\overline{X}_1 - \overline{X}_2$ is

$$V(\overline{X}_1 - \overline{X}_2) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

It seems reasonable to combine the two sample variances S_1^2 and S_2^2 to form an estimator of σ^2. The *pooled estimator* of σ^2 is defined as follows.

Pooled Estimator of Variance

The **pooled estimator** of σ^2, denoted by S_p^2, is defined by

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \tag{10.12}$$

It is easy to see that the pooled estimator S_p^2 can be written as

$$S_p^2 = \frac{n_1 - 1}{n_1 + n_2 - 2} S_1^2 + \frac{n_2 - 1}{n_1 + n_2 - 2} S_2^2 = wS_1^2 + (1 - w)S_2^2$$

where $0 < w \leq 1$. Thus, S_p^2 is a *weighted average* of the two sample variances S_1^2 and S_2^2 where the weights w and $1 - w$ depend on the two sample sizes n_1 and n_2. Obviously, if $n_1 = n_2 = n$, $w = 0.5$, S_p^2 is just the arithmetic average of S_1^2 and S_2^2. If $n_1 = 10$ and $n_2 = 20$ (say), $w = 0.32$ and $1 - w = 0.68$. The first sample contributes $n_1 - 1$ degrees of freedom to S_p^2 and the second sample contributes $n_2 - 1$ degrees of freedom. Therefore, S_p^2 has $n_1 + n_2 - 2$ degrees of freedom.

Now we know that

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

has a $N(0, 1)$ distribution. Replacing σ by S_p gives the following.

Given the assumptions of this section, the quantity

$$T = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{10.13}$$

has a t distribution with $n_1 + n_2 - 2$ degrees of freedom.

The use of this information to test the hypotheses in Equation 10.11 is now straightforward: Simply replace $\mu_1 - \mu_2$ by Δ_0, and the resulting *test statistic* has a t distribution with $n_1 + n_2 - 2$ degrees of freedom under $H_0: \mu_1 - \mu_2 = \Delta_0$. Therefore, the reference distribution for the test statistic is the t distribution with $n_1 + n_2 - 2$ degrees of freedom. The calculation of P-values and the location of the critical region for fixed-significance-level testing for both two- and one-sided alternatives parallels those in the one-sample case. Because a pooled estimate of variance is used, the procedure is often called the **pooled *t*-test**.

Tests on the Difference in Means of Two Normal Distributions, Variances Unknown and Equal*

Null hypothesis: $H_0: \mu_1 - \mu_2 = \Delta_0$

Test statistic:
$$T_0 = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{10.14}$$

*While we have given the development of this procedure for the case in which the sample sizes could be different, there is an advantage to using equal sample sizes $n_1 = n_2 = n$. When the sample sizes are the same from both populations, the t-test is more robust to the assumption of equal variances.

Alternative Hypotheses	*P*-Value	Rejection Criterion for Fixed-Level Tests
$H_1: \mu_1 - \mu_2 \neq \Delta_0$	Probability above $\lvert t_0 \rvert$ and probability below $-\lvert t_0 \rvert$	$t_0 > t_{\alpha/2,n_1+n_2-2}$ or $t_0 < -t_{\alpha/2,n_1+n_2-2}$
$H_1: \mu_1 - \mu_2 > \Delta_0$	Probability above t_0	$t_0 > t_{\alpha,n_1+n_2-2}$
$H_1: \mu_1 - \mu_2 < \Delta_0$	Probability below t_0	$t_0 < -t_{\alpha,n_1+n_2-2}$

EXAMPLE 10.5 | Yield from a Catalyst

Two catalysts are being analyzed to determine how they affect the mean yield of a chemical process. Specifically, catalyst 1 is currently used; but catalyst 2 is acceptable. Because catalyst 2 is cheaper, it should be adopted, if it does not change the process yield. A test is run in the pilot plant and results in the data shown in Table 10.1. Figure 10.2 presents a normal probability plot and a comparative box plot of the data from the two samples. Is there any difference in the mean yields? Use $\alpha = 0.05$, and assume equal variances.

TABLE 10.1 Catalyst Yield Data, Example 10.5

Observation Number	Catalyst 1	Catalyst 2
1	91.50	89.19
2	94.18	90.95
3	92.18	90.46
4	95.39	93.21
5	91.79	97.19
6	89.07	97.04
7	94.72	91.07
8	89.21	92.75
	$\bar{x}_1 = 92.255$	$\bar{x}_2 = 92.733$
	$s_1 = 2.39$	$s_2 = 2.98$

The solution using the seven-step hypothesis-testing procedure is as follows:

1. **Parameter of interest:** The parameters of interest are μ_1 and μ_2, the mean process yield using catalysts 1 and 2, respectively, and we want to know if $\mu_1 - \mu_2 = 0$.
2. **Null hypothesis:** $H_0: \mu_1 - \mu_2 = 0$, or $H_0: \mu_1 = \mu_2$
3. **Alternative hypothesis:** $H_1: \mu_1 \neq \mu_2$
4. **Test statistic:** The test statistic is

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5. **Reject H_0 if:** Reject H_0 if the *P*-value is less than 0.05.
6. **Computations:** From Table 10.1, we have $\bar{x}_1 = 92.255$, $s_1 = 2.39$, $n_1 = 8$, $\bar{x}_2 = 92.733$, $s_2 = 2.98$, and $n_2 = 8$. Therefore

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(7)(2.39)^2 + 7(2.98)^2}{8 + 8 - 2} = 7.30$$

$$s_p = \sqrt{7.30} = 2.70$$

and

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{2.70\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{92.255 - 92.733}{2.70\sqrt{\frac{1}{8} + \frac{1}{8}}} = -0.35$$

7. **Conclusions:** Because $\lvert t_0 \rvert = 0.35$, we find from Appendix Table V that $t_{0.40,14} = 0.258$ and $t_{0.25,14} = 0.692$. Therefore, because $0.258 < 0.35 < 0.692$, we conclude that lower and upper bounds on the *P*-value are $0.50 < P < 0.80$. Therefore, because the *P*-value exceeds $\alpha = 0.05$, the null hypothesis cannot be rejected.

Practical Interpretation: At the 0.05 level of significance, we do not have strong evidence to conclude that catalyst 2 results in a mean yield that differs from the mean yield when catalyst 1 is used.

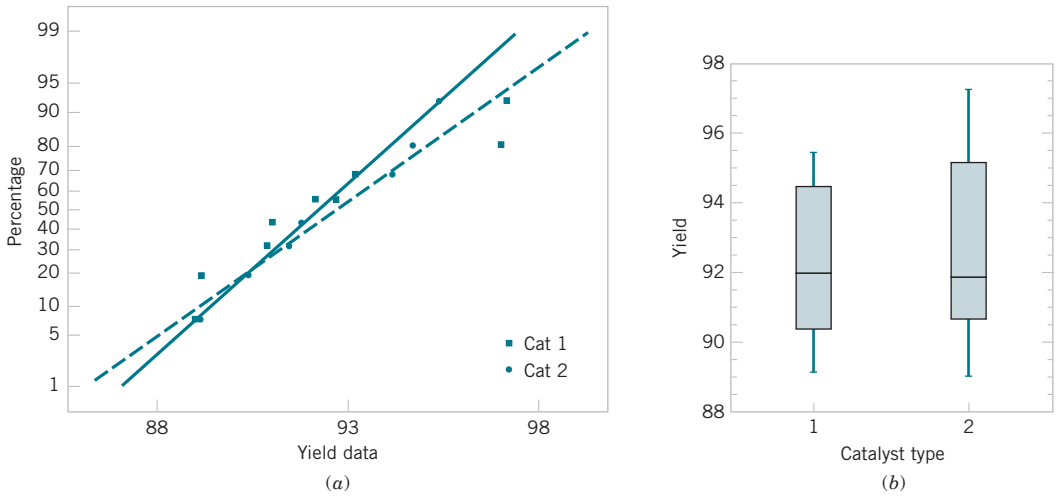


FIGURE 10.2

Normal probability plot and comparative box plot for the catalyst yield data in Example 10.5. (a) Normal probability plot. (b) Box plots.

Typical computer output for the two-sample t-test and confidence interval procedure for Example 10.5 follows:

Two-Sample T-Test and CI for Catalyst 1 vs. Catalyst 2

	N	Mean	StDev	SE Mean
Cat 1	8	92.26	2.39	0.84
Cat 2	8	92.73	2.99	1.1

```
Difference = mu Cat 1 − mu Cat 2
Estimate for difference: −0.48
95% CI for difference: (−3.37, 2.42)
T-test of difference = 0 (vs not=): T-value = −0.35 P-value = 0.730 DF = 14
Pooled StDev = 2.70
```

Notice that the numerical results are essentially the same as the manual computations in Example 10.5. The P-value is reported as $P = 0.73$. The two-sided CI on $\mu_1 - \mu_2$ is also reported. We give the computing formula for the CI in Section 10.2.3. Figure 10.2 shows the normal probability plot of the two samples of yield data and comparative box plots. The normal probability plots indicate that there is no problem with the normality assumption or with the assumption of equal variances. Furthermore, both straight lines have similar slopes, providing some verification of the assumption of equal variances. The comparative box plots indicate that there is no obvious difference in the two catalysts although catalyst 2 has slightly more sample variability.

Case 2: $\sigma_1^2 \neq \sigma_2^2$ In some situations, we cannot reasonably assume that the unknown variances σ_1^2 and σ_2^2 are equal. There is not an exact t-statistic available for testing H_0: $\mu_1 - \mu_2 = \Delta_0$ in this case. However, an approximate result can be applied.

Case 2: Test Statistic for the Difference in Means, Variances Unknown and Not Assumed Equal

If $H_0: \mu_1 - \mu_2 = \Delta_0$ is true, the statistic

$$T_0^* = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \tag{10.15}$$

is distributed approximately as t with degrees of freedom given by

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} \tag{10.16}$$

If v is not an integer, round down to the nearest integer.

Therefore, if $\sigma_1^2 \neq \sigma_2^2$, the hypotheses on differences in the means of two normal distributions are tested as in the equal variances case except that T_0^* is used as the test statistic and $n_1 + n_2 - 2$ is replaced by v in determining the degrees of freedom for the test.

The pooled t-test is very sensitive to the assumption of equal variances (so is the CI procedure in Section 10.2.3). The two-sample t-test assuming that $\sigma_1^2 \neq \sigma_2^2$ is a safer procedure unless one is very sure about the equal variance assumption.

EXAMPLE 10.6 | Arsenic in Drinking Water

Arsenic concentration in public drinking water supplies is a potential health risk. An article in the *Arizona Republic* (May 27, 2001) reported drinking water arsenic concentrations in parts per billion (ppb) for 10 metropolitan Phoenix communities and 10 communities in rural Arizona. The data follow:

Metro Phoenix ($\bar{x}_1 = 12.5,\ s_1 = 7.63$)	**Rural Arizona** ($\bar{x}_2 = 27.5,\ s_2 = 15.3$)
Phoenix, 3	Rimrock, 48
Chandler, 7	Goodyear, 44
Gilbert, 25	New River, 40
Glendale, 10	Apache Junction, 38
Mesa, 15	Buckeye, 33
Paradise Valley, 6	Nogales, 21
Peoria, 12	Black Canyon City, 20
Scottsdale, 25	Sedona, 12
Tempe, 15	Payson, 1
Sun City, 7	Casa Grande, 18

We wish to determine whether any difference exists in mean arsenic concentrations for metropolitan Phoenix communities and for communities in rural Arizona. Figure 10.3 shows a normal probability plot for the two samples of arsenic concentration. The assumption of normality appears quite reasonable, but because the slopes of the two straight lines are very different, it is unlikely that the population variances are the same.

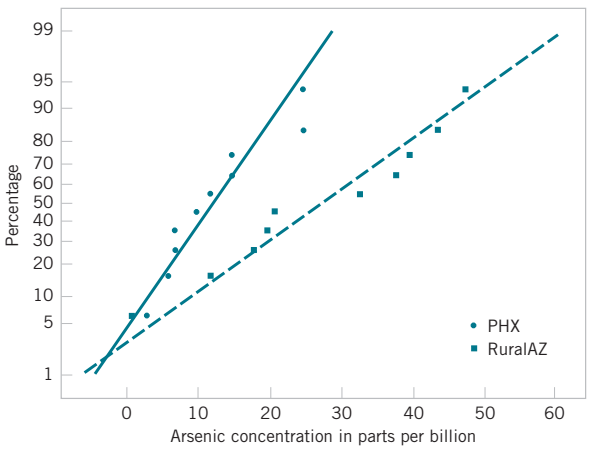


FIGURE 10.3

Normal probability plot of the arsenic concentration data.

Applying the seven-step procedure gives the following:

1. **Parameter of interest:** The parameters of interest are the mean arsenic concentrations for the two geographic regions, say, μ_1 and μ_2, and we are interested in determining whether $\mu_1 - \mu_2 = 0$.
2. **Null hypothesis:** $H_0: \mu_1 - \mu_2 = 0$, or $H_0: \mu_1 - \mu_2$
3. **Alternative hypothesis:** $H_1: \mu_1 \neq \mu_2$
4. **Test statistic:** The test statistic is
$$t_0^* = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$
5. **Reject H_0 if:** The degrees of freedom on t_0^* are found from Equation 10.16 as
$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_1)^2}{n_2 - 1}} = \frac{\left[\frac{(7.63)^2}{10} + \frac{(15.3)^2}{10}\right]^2}{\frac{[(7.63)^2/10]^2}{9} + \frac{[(15.3)^2/10]^2}{9}}$$
$$= 13.2 \simeq 13$$
Therefore, using $\alpha = 0.05$ and a fixed-significance-level test, we would reject $H_0: \mu_1 = \mu_2$ if $t_0^* > t_{0.025,13} = 2.160$ or if $t_0^* < -t_{0.025,13} = -2.160$.
6. **Computations:** Using the sample data, we find
$$t_0^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{12.5 - 27.5}{\sqrt{\frac{(7.63)^2}{10} + \frac{(15.3)^2}{10}}} = -2.77$$
7. **Conclusion:** Because $t_0^* = -2.77 < t_{0.025,13} = -2.160$, we reject the null hypothesis.

Practical Interpretation: There is strong evidence to conclude that mean arsenic concentration in the drinking water in rural Arizona is different from the mean arsenic concentration in metropolitan Phoenix drinking water. Furthermore, the mean arsenic concentration is higher in rural Arizona communities. The *P*-value for this test is approximately $P = 0.016$.

Typical computer output for this example follows:

Two-Sample T-Test and CI: PHX vs RuralAZ

	N	Mean	StDev
PHX	10	12.50	7.63
RuralAZ	10	27.50	15.3

```
Difference = mu PHX − mu RuralAZ
Estimate for difference: −15.00
95% CI for difference: (−26.71, −3.29)
T-test of difference = 0 (vs. not = ): T-value = −2.77 P-value = 0.016 DF = 13
```

The computer-generated numerical results exactly match the calculations from Example 10.6. Note that a two-sided 95% CI on $\mu_1 - \mu_2$ is also reported. We discuss its computation in Section 10.2.3; however, note that the interval does not include zero. Indeed, the upper 95% of confidence limit is −3.29 ppb, well below zero, and the mean observed difference is $\bar{x}_1 - \bar{x}_2 = 12.5 - 27.5 = -15$ ppb.

EXAMPLE 10.7 | Chocolate and Cardiovascular Health

An article in *Nature* (2003, Vol. 48, p. 1013) described an experiment in which subjects consumed different types of chocolate to determine the effect of eating chocolate on a measure of cardiovascular health. We consider the results for only dark chocolate and milk chocolate. In the experiment, 12 subjects consumed 100 grams of dark chocolate and 200 grams of milk chocolate, one type of chocolate per day, and after one hour, the total antioxidant capacity of their blood plasma was measured in an assay. The subjects consisted of seven women and five men with an average age range of 32.2 ± 1 years, an average weight of 65.8 ± 3.1 kg, and average body mass index of 21.9 ± 0.4 kg/m^2. Data similar to that reported in the article follows.

Dark Chocolate	Milk Chocolate
118.8, 122.6, 115.6, 113.6,	102.1, 105.8, 99.6, 102.7,
119.5, 115.9, 115.8, 115.1,	98.8, 100.9, 102.8, 98.7,
116.9, 115.4, 115.6, 107.9	94.7, 97.8, 99.7, 98.6

Is there evidence to support the claim that consuming dark chocolate produces a higher mean level of total blood plasma antioxidant capacity than consuming milk chocolate? Let μ_1 be the mean blood plasma antioxidant capacity resulting from eating dark chocolate and μ_2 be the mean blood plasma antioxidant capacity resulting from eating milk chocolate. The hypotheses that we wish to test are

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 > \mu_2$$

The results of applying the pooled t-test to this experiment are as follows:

Two-sample *T* for Dark vs. Milk

	N	Sample Mean	StDev
Dark	12	116.06	3.53
Milk	12	100.19	2.89

```
Difference = mu (Dark) − mu (Milk) = 15.87
T-test of difference in means = 0 (vs. mu (Dark)
> mu (Milk)):
T-value = 12.05 P-Value < 0.001
DF = 22 Pooled StDev = 3.2257
```

Because the P-value is so small (<0.001), the null hypothesis would be rejected. Strong evidence supports the claim that consuming dark chocolate produces a higher mean level of total blood plasma antioxidant capacity than consuming milk chocolate.

10.2.2 Type II Error and Choice of Sample Size

The operating characteristic curves in Appendix Charts VII*e*, VII*f*, VII*g*, and VII*h* are used to evaluate the type II error for the case in which $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Unfortunately, when $\sigma_1^2 \neq \sigma_2^2$, the distribution of T_0^* is unknown if the null hypothesis is false, and no operating characteristic curves are available for this case.

For the two-sided alternative $H_1: \mu_1 - \mu_2 = \Delta \neq \Delta_0$, when $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and $n_1 = n_2 = n$, Charts VII*e* and VII*f* are used with

$$d = \frac{|\Delta - \Delta_0|}{2\sigma} \tag{10.17}$$

where Δ is the true difference in means that is of interest. To use these curves, they must be entered with the sample size $n^* = 2n - 1$. For the one-sided alternative hypothesis, we use Charts VII*g* and VII*h* and define d and Δ as in Equation 10.17. It is noted that the parameter d is a function of σ, which is unknown. As in the single-sample t-test, we may have to rely on a prior estimate of σ or use a subjective estimate. Alternatively, we could define the differences in the mean that we wish to detect relative to σ.

EXAMPLE 10.8 | Yield from Catalyst Sample Size

Consider the catalyst experiment in Example 10.5. Suppose that, if catalyst 2 produces a mean yield that differs from the mean yield of catalyst 1 by 4.0%, we would like to reject the null hypothesis with probability at least 0.85. What sample size is required?

Using $s_p = 2.70$ as a rough estimate of the common standard deviation σ, we have $d = |\Delta|/2\sigma = |4.0|/[(2)(2.70)] = 0.74$. From Appendix Chart VII*e* with $d = 0.74$ and $\beta = 0.15$, we find $n^* = 20$, approximately. Therefore, because $n^* = 2n - 1$,

$$n = \frac{n^* + 1}{2} = \frac{20 + 1}{2} = 10.5 \simeq 11 \text{ (say)}$$

and we would use sample sizes of $n_1 = n_2 = n = 11$.

Many software packages perform power and sample size calculations for the two-sample t-test (equal variances). Typical output from Example 10.8 is as follows:

Power and Sample Size

```
2-Sample T-test
Testing mean 1 = mean 2 (versus not =)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05 Sigma = 2.7
```

Difference	Sample Size	Target Power	Actual Power
4	10	0.8500	0.8793

The results agree fairly closely with the results obtained from the OC curve.

10.2.3 Confidence Interval on the Difference in Means, Variances Unknown

Case 1: $\sigma_1^2 = \sigma_2^2 = \sigma^2$ To develop the confidence interval for the difference in means $\mu_1 - \mu_2$ when both variances are equal, note that the distribution of the statistic

$$T = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{10.18}$$

is the t distribution with $n_1 + n_2 - 2$ degrees of freedom. Therefore $P\left(-t_{\alpha/2,n_1+n_2-2} \le T \le t_{\alpha/2,n_1+n_2-2}\right) = 1 - \alpha$. Now substituting Equation 10.18 for T and manipulating the quantities inside the probability statement leads to the $100(1-\alpha)\%$ confidence interval on $\mu_1 - \mu_2$.

Case 1: Confidence Interval on the Difference in Means, Variances Unknown and Equal

If $\bar{x}_1, \bar{x}_2, s_1^2$, and s_2^2 are the sample means and variances of two random samples of sizes n_1 and n_2, respectively, from two independent normal populations with unknown but equal variances, a $100(1-\alpha)\%$ confidence interval on the difference in means $\mu_1 - \mu_2$ is

$$\bar{x}_1 - \bar{x}_2 - t_{\alpha/2,n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le \bar{x}_1 - \bar{x}_2 + t_{\alpha/2,n_1+n_2-2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{10.19}$$

where $s_p = \sqrt{[(n_1-1)s_1^2 + (n_2-1)s_2^2]/(n_1+n_2-2)}$ is the pooled estimate of the common population standard deviation, and $t_{\alpha/2,n_1+n_2-2}$ is the upper $\alpha/2$ percentage point of the t distribution with $n_1 + n_2 - 2$ degrees of freedom.

EXAMPLE 10.9 | Cement Hydration

An article in the journal *Hazardous Waste and Hazardous Materials* (1989, Vol. 6) reported the results of an analysis of the weight of calcium in standard cement and cement doped with lead. Reduced levels of calcium would indicate that the hydration mechanism in the cement is blocked and would allow water to attack various locations in the cement structure. Ten samples of standard cement had an average weight percent calcium of $\bar{x}_1 = 90.0$ with a sample standard deviation of $s_1 = 5.0$, and 15 samples of the lead-doped cement had an average weight percent calcium of $\bar{x}_2 = 87.0$ with a sample standard deviation of $s_2 = 4.0$.

We assume that weight percent calcium is normally distributed and find a 95% confidence interval on the difference in means, $\mu_1 - \mu_2$, for the two types of cement. Furthermore, we assume that both normal populations have the same standard deviation.

The pooled estimate of the common standard deviation is found using Equation 10.12 as follows:

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{9(5.0)^2 + 14(4.0)^2}{10+15-2} = 19.52$$

Therefore, the pooled standard deviation estimate is $s_p = \sqrt{19.52} = 4.4$. The 95% confidence interval is found using Equation 10.19:

$$\bar{x}_1 - \bar{x}_2 - t_{0.025,23}\, s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1 - \mu_2 \le \bar{x}_1 - \bar{x}_2 + t_{0.025,23}\, s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

or upon substituting the sample values and using $t_{0.025,23} = 2.069$,

$$90.0 - 87.0 - 2.069(4.4)\sqrt{\frac{1}{10}+\frac{1}{15}} \le \mu_1 - \mu_2 \le 90.0 - 87.0 + 2.069(4.4)\sqrt{\frac{1}{10}+\frac{1}{15}}$$

which reduces to

$$-0.72 \le \mu_1 - \mu_2 \le 6.72$$

Practical Interpretation: Notice that the 95% confidence interval includes zero; therefore, at this level of confidence we cannot conclude that there is a difference in the means. Put another way, there is no evidence that doping the cement with lead affected the mean weight percent of calcium; therefore, we cannot claim that the presence of lead affects this aspect of the hydration mechanism at the 95% level of confidence.

Case 2: $\sigma_1^2 \neq \sigma_2^2$ In many situations, assuming that $\sigma_1^2 = \sigma_2^2$ is not reasonable. When this assumption is unwarranted, we may still find a $100(1-\alpha)\%$ confidence interval on $\mu_1 - \mu_2$ using the fact that $T^* = [\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)]/\sqrt{S_1^2/n_1 + S_2^2/n_2}$ is distributed approximately as t with degrees of freedom v given by Equation 10.16. The CI expression follows.

Case 2: Approximate Confidence Interval on the Difference in Means, Variances Unknown and Not Assumed Equal

If $\bar{x}_1, \bar{x}_2, s_1^2$, and s_2^2 are the means and variances of two random samples of sizes n_1 and n_2, respectively, from two independent normal populations with unknown and unequal variances, an approximate $100(1-\alpha)\%$ confidence interval on the difference in means $\mu_1 - \mu_2$ is

$$\bar{x}_1 - \bar{x}_2 - t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \le \mu_1 - \mu_2 \le \bar{x}_1 - \bar{x}_2 + t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \tag{10.20}$$

where v is given by Equation 10.16 and $t_{\alpha/2,v}$ is the upper $\alpha/2$ percentage point of the t distribution with v degrees of freedom.

10.3 A Nonparametric Test for the Difference in Two Means

Suppose that we have two independent continuous populations X_1 and X_2 with means μ_1 and μ_2, but we are unwilling to assume that they are (approximately) normal. However, we can assume that the distributions of X_1 and X_2 are continuous and have the same shape and spread, and differ only (possibly) in their locations. The **Wilcoxon rank-sum test** can be used to test the hypothesis H_0: $\mu_1 = \mu_2$. This procedure is sometimes called the Mann-Whitney test, although the Mann-Whitney test statistic is usually expressed in a different form.

10.3.1 Description of the Wilcoxon Rank-Sum Test

Let $X_{11}, X_{12}, \ldots, X_{1n_1}$ and $X_{21}, X_{22}, \ldots, X_{2n_2}$ be two independent random samples of sizes $n_1 \leq n_2$ from the continuous populations X_1 and X_2 described earlier. We wish to test the hypotheses

$$H_0: \mu_1 = \mu_2 \qquad H_1: \mu_1 \neq \mu_2$$

The test procedure is as follows. Arrange all $n_1 + n_2$ observations in ascending order of magnitude and assign ranks to them. If two or more observations are tied (identical), use the mean of the ranks that would have been assigned if the observations differed.

Let W_1 be the sum of the ranks in the smaller sample (1), and define W_2 to be the sum of the ranks in the other sample. Then,

$$W_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - W_1 \tag{10.21}$$

Now if the sample means do not differ, we expect the sum of the ranks to be nearly equal for both samples after adjusting for the difference in sample size. Consequently, if the sums of the ranks differ greatly, we conclude that the means are not equal.

Appendix Table X contains the critical value of the rank sums for $\alpha = 0.05$ and $\alpha = 0.01$ assuming the preceding two-sided alternative. Refer to Appendix Table X with the appropriate sample sizes n_1 and n_2, and the critical value w_α can be obtained. The null $H_0: \mu_1 = \mu_2$ is rejected in favor of $H_1: \mu_1 < \mu_2$, if either of the observed values w_1 or w_2 is less than or equal to the tabulated critical value w_α.

The procedure can also be used for one-sided alternatives. If the alternative is $H_1: \mu_1 < \mu_2$, reject H_0 if $w_1 \leq w_\alpha$; for $H_1: \mu_1 > \mu_2$, reject H_0 if $w_2 \leq w_\alpha$. For these one-sided tests, the tabulated critical values w_α correspond to levels of significance of $\alpha = 0.025$ and $\alpha = 0.005$.

EXAMPLE 10.10 | Axial Stress

The mean axial stress in tensile members used in an aircraft structure is being studied. Two alloys are being investigated. Alloy 1 is a traditional material, and alloy 2 is a new aluminum-lithium alloy that is much lighter than the standard material. Ten specimens of each alloy type are tested, and the axial stress is measured. The sample data are assembled in Table 10.2. Using $\alpha = 0.05$, we wish to test the hypothesis that the means of the two stress distributions are identical.

TABLE 10.2 Axial Stress for Two Aluminum-Lithium Alloys

Alloy 1		Alloy 2	
3238 psi	3254 psi	3261 psi	3248 psi
3195	3229	3187	3215
3246	3225	3209	3226
3190	3217	3212	3240
3204	3241	3258	3234

We apply the seven-step hypothesis-testing procedure to this problem:

1. **Parameter of interest:** The parameters of interest are the means of the two distributions of axial stress.
2. **Null hypothesis:** $H_0: \mu_1 = \mu_2$
3. **Alternative hypothesis:** $H_1: \mu_1 \neq \mu_2$
4. **Test statistic:** We use the Wilcoxon rank-sum test statistic in Equation 10.21.

$$w_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - w_1$$

5. **Reject H_0 if:** Because $\alpha = 0.05$ and $n_1 = n_2 = 10$, Appendix Table X gives the critical value as $w_{0.05} = 78$. If either w_1 or w_2 is less than or equal to $w_{0.05} = 78$, we will reject $H_0: \mu_1 = \mu_2$.

6. **Computations:** The data from Table 10.2 are arranged in ascending order and ranked as follows:

Alloy Number	Axial Stress	Rank
2	3187 psi	1
1	3190	2
1	3195	3
1	3204	4
2	3209	5
2	3212	6
2	3215	7
1	3217	8
1	3225	9
2	3226	10
1	3229	11
2	3234	12
1	3238	13
2	3240	14
1	3241	15
1	3246	16
2	3248	17
1	3254	18
2	3258	19
2	3261	20

The sum of the ranks for alloy 1 is

$$w_1 = 2 + 3 + 4 + 8 + 9 + 11 + 13 + 15 + 16 + 18 = 99$$

and for alloy 2

$$\begin{aligned} w_2 &= \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - w_1 \\ &= \frac{(10 + 10)(10 + 10 + 1)}{2} - 99 \\ &= 111 \end{aligned}$$

7. **Conclusion:** Because neither w_1 nor w_2 is less than or equal to $w_{0.05} = 78$, we cannot reject the null hypothesis that both alloys exhibit the same mean axial stress.

Practical Interpretation: The data do not demonstrate that there is a superior alloy for this particular application.

10.3.2 Large-Sample Approximation

When both n_1 and n_2 are moderately large, say, more than eight, the distribution of w_1 can be well approximated by the normal distribution with mean

$$\mu_{W_1} = \frac{n_1(n_1 + n_2 + 1)}{2}$$

and variance

$$\sigma^2_{W_1} = \frac{n_1 n_2(n_1 + n_2 + 1)}{12}$$

Therefore, for n_1 and $n_2 > 8$, we could use

Normal Approximation for Wilcoxon Rank-Sum Test Statistic

$$Z_0 = \frac{W_1 - \mu_{W_1}}{\sigma_{W_1}} \tag{10.22}$$

as a statistic, and the appropriate critical region is $|z_0| > z_{\alpha/2}$, $z_0 > z_\alpha$, or $z_0 < -z_\alpha$, depending on whether the test is a two-tailed, upper-tailed, or lower-tailed test.

10.3.3 Comparison to the *t*-Test

In Chapter 9, we discussed the comparison of the *t*-test with the Wilcoxon signed-rank test. The results for the two-sample problem are similar to the one-sample case. That is, when the normality assumption is correct, the Wilcoxon rank-sum test is approximately 95% as efficient as the *t*-test in large samples. On the other hand, regardless of the form of the distributions, the Wilcoxon rank-sum test will always be at least 86% as efficient. The efficiency of the Wilcoxon test relative to the *t*-test is usually high if the underlying distribution has heavier tails than the normal because the behavior of the *t*-test is very dependent on the sample mean, which is quite unstable in heavy-tailed distributions.

10.4 Paired *t*-Test

A special case of the two-sample *t*-tests in Section 10.2 occurs when the observations on the two populations of interest are collected in *pairs*. Each pair of observations, say (X_{1j}, X_{2j}), is taken under homogeneous conditions, but these conditions may change from one pair to another. For example, suppose that we are interested in comparing two different types of tips for a hardness-testing machine. This machine presses the tip into a metal specimen with a known force. By measuring the depth of the depression caused by the tip, the hardness of the specimen can be determined. If several specimens were selected at random, half tested with tip 1, half tested with tip 2, and the pooled or independent *t*-test in Section 10.2 was applied, the results of the test could be erroneous. The metal specimens could have been cut from bar stock that was produced in different heats, or they might not be homogeneous in some other way that might affect hardness. Then the observed difference in mean hardness readings for the two tip types also includes hardness differences in specimens.

A more powerful experimental procedure is to collect the data in pairs—that is, to make two hardness readings on each specimen, one with each tip. The test procedure would then consist of analyzing the *differences* in hardness readings on each specimen. If there is no difference between tips, the mean of the differences should be zero. This test procedure is called the **paired *t*-test**.

Let $(X_{11}, X_{21}), (X_{12}, X_{22}), \ldots, (X_{1n}, X_{2n})$ be a set of n paired observations for which we assume that the mean and variance of the population represented by X_1 are μ_1 and σ_1^2, and the mean and variance of the population represented by X_2 are μ_2 and σ_2^2. Define the difference for each pair of observations as $D_j = X_{1j} - X_{2j}, j = 1, 2, \ldots, n$. The D_j's are assumed to be normally distributed with mean

$$\mu_D = E(X_1 - X_2) = E(X_1) - E(X_2) = \mu_1 - \mu_2$$

and variance σ_D^2, so testing hypotheses about the difference for μ_1 and μ_2 can be accomplished by performing a one-sample *t*-test on μ_D. Specifically, testing $H_0: \mu_1 - \mu_2 = \Delta_0$ against $H_1: \mu_1 - \mu_2 \neq \Delta_0$ is equivalent to testing

$$\begin{aligned} H_0&: \mu_D = \Delta_0 \\ H_1&: \mu_D \neq \Delta_0 \end{aligned} \tag{10.23}$$

The test statistic and decision procedure follow.

Paired *t*-Test

Null hypothesis: $H_0: \mu_D = \Delta_0$

Test statistic:
$$T_0 = \frac{\overline{D} - \Delta_0}{S_D/\sqrt{n}} \tag{10.24}$$

Alternative Hypotheses	*P*-Value	Rejection Criterion for Fixed-Level Tests
$H_1: \mu_D \neq \Delta_0$	Probability above $\lvert t_0 \rvert$ and probability below $-\lvert t_0 \rvert$	$t_0 > t_{\alpha/2,n-1}$ or $t_0 < -t_{\alpha/2,n-1}$
$H_1: \mu_D > \Delta_0$	Probability above t_0	$t_0 > t_{\alpha,n-1}$
$H_1: \mu_D < \Delta_0$	Probability below t_0	$t_0 < -t_{\alpha,n-1}$

In Equation 10.24, $\overline{D}$ is the sample average of the n differences $D_1, D_2, \ldots, D_n$, and S_D is the sample standard deviation of these differences.

EXAMPLE 10.11 | Shear Strength of Steel Girder

An article in the *Journal of Strain Analysis for Engineering Design* ["Model Studies on Plate Girders" (1983, Vol. 18(2), pp. 111–117)] reports a comparison of several methods for predicting the shear strength for steel plate girders. Data for two of these methods, the Karlsruhe and Lehigh procedures, when applied to nine specific girders, are shown in Table 10.3. We wish to determine whether there is any difference (on the average) for the two methods.

TABLE 10.3 **Strength Predictions for Nine Steel Plate Girders (Predicted Load/Observed Load)**

Girder	Karlsruhe Method	Lehigh Method	Difference d_j
S1/1	1.186	1.061	0.125
S2/1	1.151	0.992	0.159
S3/1	1.322	1.063	0.259
S4/1	1.339	1.062	0.277
S5/1	1.200	1.065	0.135
S2/1	1.402	1.178	0.224
S2/2	1.365	1.037	0.328
S2/3	1.537	1.086	0.451
S2/4	1.559	1.052	0.507

The seven-step procedure is applied as follows:

1. **Parameter of interest:** The parameter of interest is the difference in mean shear strength for the two methods—say, $\mu_D = \mu_1 - \mu_2 = 0$.
2. **Null hypothesis:** $H_0: \mu_D = 0$
3. **Alternative hypothesis:** $H_1: \mu_D \neq 0$
4. **Test statistic:** The test statistic is
$$t_0 = \frac{\overline{d}}{s_d/\sqrt{n}}$$
5. **Reject H_0 if:** Reject H_0 if the *P*-value <0.05.
6. **Computations:** The sample average and standard deviation of the differences d_j are $\overline{d} = 0.2739$ and $s_d = 0.1350$, and so the test statistic is
$$t_0 = \frac{\overline{d}}{s_d/\sqrt{n}} = \frac{0.2739}{0.1350/\sqrt{9}} = 6.08$$
7. **Conclusion:** Because $t_{0.0005,8} = 5.041$ and the value of the test statistic $t_0 = 6.08$ exceeds this value, the *P*-value is less than 2(0.0005) = 0.001. Therefore, we conclude that the strength prediction methods yield different results.

Practical Interpretation: Specifically, the data indicate that the Karlsruhe method produces, on the average, greater strength predictions than does the Lehigh method. This is a strong conclusion.

Software can perform the paired t-test. Typical output for Example 10.11 follows:

Paired T for Karlsruhe–Lehigh

	N	Mean	StDev	SE Mean
Karlsruhe	9	1.34011	0.14603	0.04868
Lehigh	9	1.06322	0.05041	0.01680
Difference	9	0.276889	0.135027	0.045009

95% CI for mean difference: (0.173098, 0.380680)
T-test of mean difference = 0 (vs. not = 0): T-value = 6.15, P-value = 0.000

The results essentially agree with the manual calculations, in addition to the hypothesis test results. Most computer software report a two-sided CI on the difference in means. This CI was found by constructing a single-sample CI on μ_D. We provide the details later.

Paired Versus Unpaired Comparisons In performing a comparative experiment, the investigator can sometimes choose between the paired experiment and the two-sample (or unpaired) experiment. If n measurements are to be made on each population, the two-sample t-statistic is

$$T_0 = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{S_p\sqrt{\frac{1}{n} + \frac{1}{n}}}$$

which would be compared to t_{2n-2}, and of course, the paired t-statistic is

$$T_0 = \frac{\overline{D} - \Delta_0}{S_D/\sqrt{n}}$$

which is compared to t_{n-1}. Notice that because

$$\overline{D} = \sum_{j=1}^{n} \frac{D_j}{n} = \sum_{j=1}^{n} \frac{(X_{1j} - X_{2j})}{n} = \sum_{j=1}^{n} \frac{X_{1j}}{n} - \sum_{j=1}^{n} \frac{X_{2j}}{n} = \overline{X}_1 - \overline{X}_2$$

the numerators of both statistics are identical. However, the denominator of the two-sample t-test is based on the assumption that X_1 and X_2 are *independent*. In many paired experiments, a strong positive correlation ρ exists for X_1 and X_2. Then it can be shown that

$$V(\overline{D}) = V(\overline{X}_1 - \overline{X}_2 - \Delta_0) = V(\overline{X}_1) + V(\overline{X}_2) - 2\,\mathrm{cov}(\overline{X}_1, \overline{X}_2) = \frac{2\sigma^2(1-\rho)}{n}$$

assuming that both populations X_1 and X_2 have identical variances σ^2. Furthermore, S_D^2/n estimates the variance of $\overline{D}$. Whenever a positive correlation exists within the pairs, the denominator for the paired t-test will be smaller than the denominator of the two-sample t-test. This can cause the two-sample t-test to considerably understate the significance of the data if it is incorrectly applied to paired samples.

Although pairing will often lead to a smaller value of the variance of $\overline{X}_1 - \overline{X}_2$, it does have a disadvantage—namely, the paired t-test leads to a loss of $n - 1$ degrees of freedom in comparison to the two-sample t-test. Generally, we know that increasing the degrees of freedom of a test increases the power against any fixed alternative values of the parameter.

So how do we decide to conduct the experiment? Should we pair the observations or not? Although this question has no general answer, we can give some guidelines based on the preceding discussion.

1. If the experimental units are relatively homogeneous (small σ) and the correlation within pairs is small, the gain in precision attributable to pairing will be offset by the loss of degrees of freedom, so an independent-sample experiment should be used.
2. If the experimental units are relatively heterogeneous (large σ) and there is large positive correlation within pairs, the paired experiment should be used. Typically, this case occurs when the experimental units are the *same* for both treatments; as in Example 10.11, the same girders were used to test the two methods.

Implementing the rules still requires judgment because σ and ρ are never known precisely. Furthermore, if the number of degrees of freedom is large (say, 40 or 50), the loss of $n - 1$ of them for pairing may not be serious. However, if the number of degrees of freedom is small (say, 10 or 20), losing half of them is potentially serious if not compensated for by increased precision from pairing.

Confidence Interval for μ_D To construct the confidence interval for $\mu_D = \mu_1 - \mu_2$, note that

$$T = \frac{\overline{D} - \mu_D}{S_D/\sqrt{n}}$$

follows a t distribution with $n - 1$ degrees of freedom. Then, because $P(-t_{\alpha/2,n-1} \le T \le t_{\alpha/2,n-1}) = 1 - \alpha$, we can substitute for T in the preceding expression and perform the necessary steps to isolate $\mu_D = \mu_1 - \mu_2$ for the inequalities. This leads to the following $100(1 - \alpha)\%$ confidence interval on $\mu_1 - \mu_2$.

Confidence Interval for μ_D from Paired Samples

If $\overline{d}$ and s_D are the sample mean and standard deviation of the difference of n random pairs of normally distributed measurements, a $100(1 - \alpha)\%$ confidence interval on the difference in means $\mu_D = \mu_1 - \mu_2$ is

$$\overline{d} - t_{\alpha/2,n-1}\, s_D/\sqrt{n} \le \mu_D \le \overline{d} + t_{\alpha/2,n-1}\, s_D/\sqrt{n} \tag{10.25}$$

where $t_{\alpha/2,n-1}$ is the upper $100\alpha/2$ percentage point of the t distribution with $n - 1$ degrees of freedom.

This confidence interval is also valid for the case in which $\sigma_1^2 \ne \sigma_2^2$ because s_D^2 estimates $\sigma_D^2 = V(X_1 - X_2)$. Also, for large samples (say, $n \ge 30$ pairs), the explicit assumption of normality is unnecessary because of the central limit theorem.

EXAMPLE 10.12 | Parallel Park Cars

The journal *Human Factors* ["Relative Controllability of Dissimilar Cars" (1962, Vol. 4(6), pp. 375–380)] reported a study in which $n = 14$ subjects were asked to parallel-park two cars having very different wheel bases and turning radii. The time in seconds for each subject was recorded and is given in Table 10.4. From the column of observed differences, we calculate $\overline{d} = 1.21$ and $s_D = 12.68$. The 90% confidence

TABLE 10.4 Time in Seconds to Parallel-Park Two Automobiles

	Automobile		Difference
Subject	$1(x_{1j})$	$2(x_{2j})$	(d_j)
1	37.0	17.8	19.2
2	25.8	20.2	5.6
3	16.2	16.8	−0.6
4	24.2	41.4	−17.2
5	22.0	21.4	0.6
6	33.4	38.4	−5.0
7	23.8	16.8	7.0
8	58.2	32.2	26.0
9	33.6	27.8	5.8
10	24.4	23.2	1.2
11	23.4	29.6	−6.2
12	21.2	20.6	0.6
13	36.2	32.2	4.0
14	29.8	53.8	−24.0

interval for $\mu_D = \mu_1 - \mu_2$ is found from Equation 10.25 as follows:

$$\bar{d} - t_{0.05,13}\, s_D/\sqrt{n} \le \mu_D \le \bar{d} + t_{0.05,13}\, s_D/\sqrt{n}$$
$$1.21 - 1.771(12.68)/\sqrt{14} \le \mu_D \le 1.21 + 1.771(12.68)/\sqrt{14}$$
$$-4.79 \le \mu_D \le 7.21$$

Notice that the confidence interval on μ_D includes zero. This implies that, at the 90% level of confidence, the data do not support the claim that the two cars have different mean parking times μ_1 and μ_2. That is, the value $\mu_D = \mu_1 - \mu_2 = 0$ is not inconsistent with the observed data.

Nonparametric Approach to Paired Comparisons Both the sign test and the Wilcoxon signed-rank test discussed in Section 9.9 can be applied to paired observations. In the case of the sign test, the null hypothesis is that the median of the differences is equal to zero (that is, H_0: $\tilde{\mu}_D = 0$). The Wilcoxon signed-rank test is for the null hypothesis that the mean of the differences is equal to zero. The procedures are applied to the observed differences as described in Sections 9.9.1 and 9.9.2.

10.5 Inference on the Variances of Two Normal Distributions

We now introduce tests and confidence intervals for the two population variances shown in Figure 10.1. We assume that both populations are normal. Both the hypothesis-testing and confidence interval procedures are relatively sensitive to the normality assumption.

10.5.1 *F* Distribution

Suppose that two independent normal populations are of interest when the population means and variances, say, μ_1, σ_1^2, μ_2, and σ_2^2, are unknown. We wish to test hypotheses about the equality of the two variances, say, H_0: $\sigma_1^2 = \sigma_2^2$. Assume that two random samples of size n_1 from population 1 and of size n_2 from population 2 are available, and let S_1^2 and S_2^2 be the sample variances. We wish to test the hypotheses

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \ne \sigma_2^2 \qquad (10.26)$$

The development of a test procedure for these hypotheses requires a new probability distribution, the F distribution. The random variable F is defined to be the ratio of two independent chi-square random variables, each divided by its number of degrees of freedom. That is,

$$F = \frac{W/u}{Y/v}$$

where W and Y are independent chi-square random variables with u and v degrees of freedom, respectively. We now formally state the sampling distribution of F.

Let W and Y be independent chi-square random variables with u and v degrees of freedom, respectively. Then the ratio

$$F = \frac{W/u}{Y/v} \tag{10.27}$$

has the probability density function

$$f(x) = \frac{\Gamma\left(\frac{u+v}{2}\right)\left(\frac{u}{v}\right)^{u/2} x^{(u/2)-1}}{\Gamma\left(\frac{u}{2}\right)\Gamma\left(\frac{v}{2}\right)\left[\left(\frac{u}{v}\right)x+1\right]^{(u+v)/2}}, \qquad 0 < x < \infty \tag{10.28}$$

and is said to follow the F distribution with u degrees of freedom in the numerator and v degrees of freedom in the denominator. It is usually abbreviated as $F_{u,v}$.

The mean and variance of the F distribution are $\mu = v/(v-2)$ for $v > 2$, and

$$\sigma^2 = \frac{2v^2(u+v-2)}{u(v-2)^2(v-4)}, \qquad v > 4 \tag{10.29}$$

Two F distributions are shown in Figure 10.4. The F random variable is nonnegative, and the distribution is skewed to the right. The F distribution looks very similar to the chi-square distribution; however, the two parameters u and v provide extra flexibility regarding shape.

The percentage points of the F distribution are given in Table VI of the Appendix. Let $f_{\alpha,u,v}$ be the percentage point of the F distribution with numerator degrees of freedom u and denominator degrees of freedom v such that the probability that the random variable F exceeds this value is

$$P(F > f_{\alpha,u,v}) = \int_{f_{\alpha,u,v}}^{\infty} f(x)\,dx = \alpha$$

This is illustrated in Figure 10.5. For example, if $u = 5$ and $v = 10$, we find from Table V of the Appendix that

$$P(F > f_{0.05,5,10}) = P(F_{5,10} > 3.33) = 0.05$$

That is, the upper 5 percentage points of $F_{5,10}$ is $f_{0.05,5,10} = 3.33$.

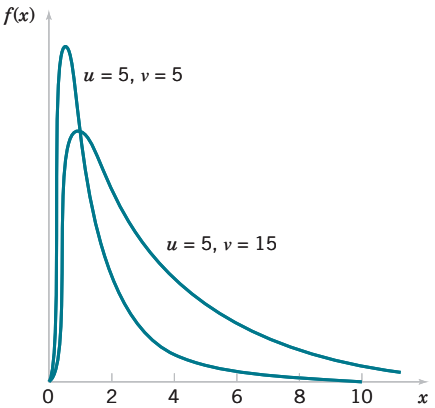


FIGURE 10.4
Probability density functions of two F distributions.

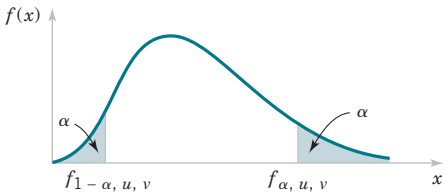


FIGURE 10.5
Upper and lower percentage points of the F distribution.

Table VI contains only upper-tailed percentage points (for selected values of $f_{\alpha,u,v}$ for $\alpha \leq 0.25$) of the F distribution. The lower-tailed percentage points $f_{1-\alpha,u,v}$ can be found as follows.

Finding Lower-Tail Points of the *F*-Distribution

$$f_{1-\alpha,u,v} = \frac{1}{f_{\alpha,v,u}} \tag{10.30}$$

For example, to find the lower-tailed percentage point $f_{0.95,5,10}$, note that

$$f_{0.95,5,10} = \frac{1}{f_{0.05,10,5}} = \frac{1}{4.74} = 0.211$$

10.5.2 Hypothesis Tests on the Equity of Two Variances

A hypothesis-testing procedure for the equality of two variances is based on the following result.

Distribution of the Ratio of Sample Variances from Two Normal Distributions

Let $X_{11}, X_{12}, \ldots, X_{1n_1}$ be a random sample from a normal population with mean μ_1 and variance σ_1^2, and let $X_{21}, X_{22}, \ldots, X_{2n_2}$ be a random sample from a second normal population with mean μ_2 and variance σ_2^2. Assume that both normal populations are independent. Let S_1^2 and S_2^2 be the sample variances. Then the ratio

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

has an F distribution with $n_1 - 1$ numerator degrees of freedom and $n_2 - 1$ denominator degrees of freedom.

This result is based on the fact that $(n_1 - 1)\, S_1^2/\sigma_1^2$ is a chi-square random variable with $n_1 - 1$ degrees of freedom, that $(n_2 - 1)\, S_2^2/\sigma_2^2$ is a chi-square random variable with $n_2 - 1$ degrees of freedom, and that the two normal populations are independent. Clearly, under the null hypothesis $H_0\colon \sigma_1^2 = \sigma_2^2$, the ratio $F_0 = S_1^2/S_2^2$ has an F_{n_1-1,n_2-1} distribution. This is the basis of the following test procedure.

Tests on the Equality of Variances from Two Normal Distributions

Null hypothesis: $H_0\colon \sigma_1^2 = \sigma_2^2$

Test statistic: $$F_0 = \frac{S_1^2}{S_2^2} \tag{10.31}$$

Alternative Hypotheses	Rejection Criterion
$H_1\colon \sigma_1^2 \neq \sigma_2^2$	$f_0 > f_{\alpha/2,n_1-1,n_2-1}$ or $f_0 < f_{1-\alpha/2,n_1-1,n_2-1}$
$H_1\colon \sigma_1^2 > \sigma_2^2$	$f_0 > f_{\alpha,n_1-1,n_2-1}$
$H_1\colon \sigma_1^2 < \sigma_2^2$	$f_0 < f_{1-\alpha,n_1-1,n_2-1}$

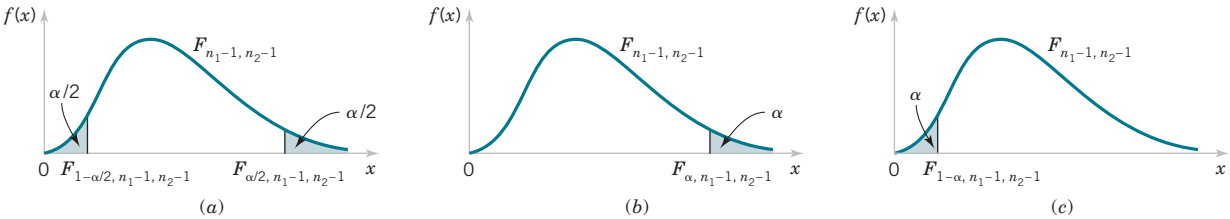


FIGURE 10.6

The F distribution for the test of H_0: $\sigma_1^2 = \sigma_2^2$ with critical region values for (a) H_1: $\sigma_1^2 \neq \sigma_2^2$, (b) H_1: $\sigma_1^2 > \sigma_2^2$, and (c) H_1: $\sigma_1^2 < \sigma_2^2$.

The critical regions for these fixed-significance-level tests are shown in Figure 10.6. Remember that this procedure is relatively sensitive to the normality assumption.

EXAMPLE 10.13 | Semiconductor Etch Variability

Oxide layers on semiconductor wafers are etched in a mixture of gases to achieve the proper thickness. The variability in the thickness of these oxide layers is a critical characteristic of the wafer, and low variability is desirable for subsequent processing steps. Two different mixtures of gases are being studied to determine whether one is superior in reducing the variability of the oxide thickness. Sixteen wafers are etched in each gas. The sample standard deviations of oxide thickness are $s_1 = 1.96$ angstroms and $s_2 = 2.13$ angstroms, respectively. Is there any evidence to indicate that either gas is preferable? Use a fixed-level test with $\alpha = 0.05$.

The seven-step hypothesis-testing procedure may be applied to this problem as follows:

1. **Parameter of interest:** The parameters of interest are the variances of oxide thickness σ_1^2 and σ_2^2. We assume that oxide thickness is a normal random variable for both gas mixtures.
2. **Null hypothesis:** $H_0: \sigma_1^2 = \sigma_2^2$
3. **Alternative hypothesis:** $H_1: \sigma_1^2 \neq \sigma_2^2$
4. **Test statistic:** The test statistic is given by Equation 10.31:
$$f_0 = \frac{s_1^2}{s_2^2}$$
5. **Reject H_0 if:** Because $n_1 = n_2 = 16$ and $\alpha = 0.05$, we will reject $H_0: \sigma_1^2 = \sigma_2^2$ if $f_0 > f_{0.025,15,15} = 2.86$ or if $f_0 < f_{0.975,15,15} = 1/f_{0.025,15,15} = 1/2.86 = 0.35$. Refer to Figure 10.6(a).
6. **Computations:** Because $s_1^2 = (1.96)^2 = 3.84$ and $s_2^2 = (2.13)^2 = 4.54$, the test statistic is
$$f_0 = \frac{s_1^2}{s_2^2} = \frac{3.84}{4.54} = 0.85$$
7. **Conclusion:** Because $f_{0.975,15,15} = 0.35 < 0.85 < f_{0.025,15,15} = 2.86$, we cannot reject the null hypothesis $H_0: \sigma_1^2 = \sigma_2^2$ at the 0.05 level of significance.

Practical Interpretation: There is no strong evidence to indicate that either gas results in a smaller variance of oxide thickness.

P-Values for the F-Test The P-value approach can also be used with F-tests. To show how to do this, consider the upper-tailed test. The P-value is the area (probability) under the F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom that lies beyond the computed value of the test statistic f_0. Appendix A Table IV can be used to obtain upper and lower bounds on the P-value. For example, consider an F-test with 9 numerator and 14 denominator degrees of freedom for which $f_0 = 3.05$. From Appendix A Table IV, we find that $f_{0.05,9,14} = 2.65$ and $f_{0.025,9,14} = 3.21$, so because $f_0 = 3.05$ lies between these two values, the P-value is between 0.05 and 0.025; that is, $0.025 < P < 0.05$. The P-value for a lower-tailed test would be found similarly, although Appendix A Table IV contains only upper-tailed points of the F distribution, Equation 10.31 would have to be used to find the necessary lower-tail points. For a two-tailed test, the bounds obtained from a one-tailed test would be doubled to obtain the P-value.

To illustrate calculating bounds on the P-value for a two-tailed F-test, reconsider Example 10.13. The computed value of the test statistic in this example is $f_0 = 0.85$. This value

falls in the lower tail of the $F_{15,15}$ distribution. The lower-tailed point that has 0.25 probability to the left of it is $f_{0.75,15,15} = 1/f_{0.25,15,15} = 1/1.43 = 0.70$, and because $0.70 < 0.85$, the probability that lies to the left of 0.85 exceeds 0.25. Therefore, we would conclude that the P-value for $f_0 = 0.85$ is greater than $2(0.25) = 0.5$, so there is insufficient evidence to reject the null hypothesis. This is consistent with the original conclusions from Example 10.13. The actual P-value is 0.7570. This value was obtained from a calculator from which we found that $P(F_{15,15} \le 0.85) = 0.3785$ and $2(0.3785) = 0.7570$. Computer software can also be used to calculate the required probabilities.

Some computer packages will perform the F-test on the equality of two variances of independent normal distributions. The output from the computer package follows.

Test for Equal Variances

```
95% Bonferroni confidence intervals for standard deviations
Sample     N      Lower      StDev      Upper
   1      16    1.38928    1.95959    3.24891
   2      16    1.51061    2.13073    3.53265
F-test (normal distribution)
Test statistic = 0.85, P-value = 0.750
```

Computer software also gives confidence intervals on the individual variances. These are the confidence intervals originally given in Equation 8.19 except that a Bonferroni "adjustment" has been applied to make the confidence level for both intervals simultaneously equal to at least 95%. This consists of using $\alpha/2 = 0.05/2 = 0.025$ to construct the individual intervals. That is, each individual confidence interval is a 97.5% CI. In Section 10.5.4, we show how to construct a CI on the *ratio* of the two variances.

10.5.3 Type II Error and Choice of Sample Size

Appendix Charts VIIo, VIIp, VIIq, and VIIr provide operating characteristic curves for the F-test given in Section 10.5.1 for $\alpha = 0.05$ and $\alpha = 0.01$, assuming that $n_1 = n_2 = n$. Charts VIIo and VIIp are used with the two-sided alternate hypothesis. They plot β against the abscissa parameter

$$\lambda = \frac{\sigma_1}{\sigma_2} \tag{10.32}$$

for various $n_1 = n_2 = n$. Charts VIIq and VIIr are used for the one-sided alternative hypotheses.

EXAMPLE 10.14 | Semiconductor Etch Variability Sample Size

For the semiconductor wafer oxide etching problem in Example 10.13, suppose that one gas resulted in a standard deviation of oxide thickness that is half the standard deviation of oxide thickness of the other gas. If we wish to detect such a situation with probability at least 0.80, is the sample size $n_1 = n_2 = 20$ adequate?

Note that if one standard deviation is half the other,

$$\lambda = \frac{\sigma_1}{\sigma_2} = 2$$

By referring to Appendix Chart VIIo with $n_1 = n_2 = 20$ and $\lambda = 2$, we find that $\beta \simeq 0.20$. Therefore, if $\beta \simeq 0.20$, the power of the test (which is the probability that the difference in standard deviations will be detected by the test) is 0.80, and we conclude that the sample sizes $n_1 = n_2 = 20$ are adequate.

10.5.4 Confidence Interval on the Ratio of Two Variances

To find the confidence interval on σ_1^2/σ_2^2, recall that the sampling distribution of

$$F = \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2}$$

is an F with $n_2 - 1$ and $n_1 - 1$ degrees of freedom. Therefore, $P(f_{1-\alpha/2,n_2-1,n_1-1} \le F \le f_{\alpha/2,n_2-1,n_1-1}) = 1 - \alpha$. Substitution for F and manipulation of the inequalities will lead to the $100(1 - \alpha)\%$ confidence interval for σ_1^2/σ_2^2.

Confidence Interval on the Ratio of Variances from Two Normal Distributions

If s_1^2 and s_2^2 are the sample variances of random samples of sizes n_1 and n_2, respectively, from two independent normal populations with unknown variances σ_1^2 and σ_2^2, then a $100(1 - \alpha)\%$ confidence interval on the ratio σ_1^2/σ_2^2 is

$$\frac{s_1^2}{s_2^2} f_{1-\alpha/2,n_2-1,n_1-1} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_1^2}{s_2^2} f_{\alpha/2,n_2-1,n_1-1} \tag{10.33}$$

where $f_{\alpha/2,n_2-1,n_1-1}$ and $f_{1-\alpha/2,n_2-1,n_1-1}$ are the upper and lower $\alpha/2$ percentage points of the F distribution with $n_2 - 1$ numerator and $n_1 - 1$ denominator degrees of freedom, respectively. A confidence interval on the ratio of the standard deviations can be obtained by taking square roots in Equation 10.33.

Just as in the hypothesis testing procedure, this CI is relatively sensitive to the normality assumption.

EXAMPLE 10.15 | Surface Finish for Titanium Alloy

A company manufactures impellers for use in jet-turbine engines. One of the operations involves grinding a particular surface finish on a titanium alloy component. Two different grinding processes can be used, and both processes can produce parts at identical mean surface roughness. The manufacturing engineer would like to select the process having the least variability in surface roughness. A random sample of $n_1 = 11$ parts from the first process results in a sample standard deviation $s_1 = 5.1$ microinches, and a random sample of $n_2 = 16$ parts from the second process results in a sample standard deviation of $s_2 = 4.7$ microinches. We will find a 90% confidence interval on the ratio of the two standard deviations, σ_1/σ_2.

Assuming that the two processes are independent and that surface roughness is normally distributed, we can use Equation 10.33 as follows:

$$\frac{s_1^2}{s_2^2} f_{0.95,15,10} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_1^2}{s_2^2} f_{0.05,15,10}$$

$$\frac{(5.1)^2}{(4.7)^2}\,0.39 \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{(5.1)^2}{(4.7)^2}\,2.85$$

or upon completing the implied calculations and taking square roots,

$$0.678 \le \frac{\sigma_1}{\sigma_2} \le 1.832$$

Notice that we have used Equation 10.30 to find $f_{0.95,15,10} = 1/f_{0.05,10,15} = 1/2.54 = 0.39$.

Practical Interpretation: Because this confidence interval includes unity, we cannot claim that the standard deviations of surface roughness for the two processes are different at the 90% level of confidence.

10.6 Inference on Two Population Proportions

We now consider the case with two binomial parameters of interest, say, p_1 and p_2, and we wish to draw inferences about these proportions. We present large-sample hypothesis testing and confidence interval procedures based on the normal approximation to the binomial.

10.6.1 Large-Sample Tests on the Difference in Population Proportions

Suppose that two independent random samples of sizes n_1 and n_2 are taken from two populations, and let X_1 and X_2 represent the number of observations that belong to the class of interest in samples 1 and 2, respectively. Furthermore, suppose that the normal approximation to the binomial is applied to each population, so the estimators of the population proportions $P_1 = X_1/n_1$ and $P_2 = X_2/n_2$ have approximate normal distributions. We are interested in testing the hypotheses

$$H_0: p_1 = p_2 \qquad H_1: p_1 \neq p_2$$

The statistic

Test Statistic for the Difference of Two Population Proportions

$$Z = \frac{\hat{P}_1 - \hat{P}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \tag{10.34}$$

is distributed approximately as standard normal and is the basis of a test for $H_0: p_1 = p_2$. Specifically, if the null hypothesis $H_0: p_1 = p_2$ is true, by using the fact that $p_1 = p_2 = p$, the random variable

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

is distributed approximately $N(0, 1)$. A pooled estimator of the common parameter p is

$$\hat{P} = \frac{X_1 + X_2}{n_1 + n_2}$$

The test statistic for $H_0: p_1 = p_2$ is then

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

This leads to the test procedures described as follows.

Approximate Tests on the Difference of Two Population Proportions

Null hypothesis: $H_0: p_1 = p_2$

Test statistic:
$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{10.35}$$

Alternative Hypotheses	*P*-Value	Rejection Criterion for Fixed-Level Tests
$H_1: p_1 \neq p_2$	Probability above $\lvert z_0 \rvert$ and probability below $-\lvert z_0 \rvert$, $P = 2[1 - \Phi(\lvert z_0 \rvert)]$	$z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$
$H_1: p_1 > p_2$	Probability above z_0, $P = 1 - \Phi(z_0)$	$z_0 > z_\alpha$
$H_1: p_1 < p_2$	Probability below z_0, $P = \Phi(z_0)$	$z_0 < -z_\alpha$

EXAMPLE 10.16 | St. John's Wort

Extracts of St. John's Wort are widely used to treat depression. An article in the April 18, 2001, issue of the *Journal of the American Medical Association* ("Effectiveness of St. John's Wort on Major Depression: A Randomized Controlled Trial") compared the efficacy of a standard extract of St. John's Wort with a placebo in 200 outpatients diagnosed with major depression. Patients were randomly assigned to two groups; one group received the St. John's Wort, and the other received the placebo. After 8 weeks, 19 of the placebo-treated patients showed improvement, and 27 of those treated with St. John's Wort improved. Is there any reason to believe that St. John's Wort is effective in treating major depression? Use $\alpha = 0.05$.

The seven-step hypothesis testing procedure leads to the following results:

1. **Parameter of interest:** The parameters of interest are p_1 and p_2, the proportion of patients who improve following treatment with St. John's Wort (p_1) or the placebo (p_2).

2. **Null hypothesis:** $H_0: p_1 = p_2$

3. **Alternative hypothesis:** $H_1: p_1 > p_2$

4. **Test statistic:** The test statistic is

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $\hat{p}_1 = 27/100 = 0.27$, $\hat{p}_2 = 19/100 = 0.19$, $n_1 = n_2 = 100$, and

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{19 + 27}{100 + 100} = 0.23$$

5. **Reject H_0 if:** Reject $H_0: p_1 = p_2$ if the P-value is less than 0.05.

6. **Computation:** The value of the test statistic is

$$z_0 = \frac{0.27 - 0.19}{\sqrt{0.23(0.77)\left(\frac{1}{100} + \frac{1}{100}\right)}} = 1.34$$

7. **Conclusion:** Because $z_0 = 1.34$, the P-value is $P = [1 - \Phi(1.34)] = 0.09$, so, we cannot reject the null hypothesis.

Practical Interpretation: There is insufficient evidence to support the claim that St. John's Wort is effective in treating major depression.

The following display shows a typical computer output for the two-sample hypothesis test and CI procedure for proportions. Notice that the 95% CI on $p_1 - p_2$ includes zero. The equation for constructing the CI is given in Section 10.6.3.

Test and CI for Two Proportions

Sample	X	N	Sample p
1	27	100	0.270
2	19	100	0.190

```
Estimate for p(1) - p(2): 0.08
95% CI for p(1) - p(2): (-0.0361186, 0.196119)
Test for p(1) - p(2) = 0 (vs not = 0): Z = 1.35 P-value = 0.177
```

10.6.2 Type II Error and Choice of Sample Size

The computation of the β-error for the large-sample test of $H_0: p_1 = p_2$ is somewhat more involved than in the single-sample case. The problem is that the denominator of the test statistic Z_0 is an estimate of the standard deviation of $\hat{P}_1 - \hat{P}_2$ under the assumption that $p_1 = p_2 = p$. When $H_0: p_1 = p_2$ is false, the standard deviation of $\hat{P}_1 - \hat{P}_2$ is

$$\sigma_{\hat{P}_1 - \hat{P}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \tag{10.36}$$

Approximate Type II Error for a Two-Sided Test on the Difference of Two Population Proportions

If the alternative hypothesis is two sided, the β-error is

$$\beta = \Phi\left[\frac{z_{\alpha/2}\sqrt{\bar{p}\bar{q}(1/n_1 + 1/n_2)} - (p_1 - p_2)}{\sigma_{\hat{P}_1 - \hat{P}_2}}\right] - \Phi\left[\frac{-z_{\alpha/2}\sqrt{\bar{p}\bar{q}(1/n_1 + 1/n_2)} - (p_1 - p_2)}{\sigma_{\hat{P}_1 - \hat{P}_2}}\right] \tag{10.37}$$

where

$$\bar{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \quad \text{and} \quad \bar{q} = \frac{n_1(1-p_1) + n_2(1-p_2)}{n_1 + n_2}$$

and $\sigma_{\hat{P}_1 - \hat{P}_2}$ is given by Equation 10.36.

Approximate Type II Error for a One-Sided Test on the Difference of Two Population Proportions

If the alternative hypothesis is $H_1: p_1 > p_2$,

$$\beta = \Phi\left[\frac{z_{\alpha}\sqrt{\bar{p}\bar{q}(1/n_1 + 1/n_2)} - (p_1 - p_2)}{\sigma_{\hat{P}_1 - \hat{P}_2}}\right] \tag{10.38}$$

and if the alternative hypothesis is $H_1: p_1 < p_2$,

$$\beta = 1 - \Phi\left[\frac{-z_{\alpha}\sqrt{\bar{p}\bar{q}(1/n_1 + 1/n_2)} - (p_1 - p_2)}{\sigma_{\hat{P}_1 - \hat{P}_2}}\right] \tag{10.39}$$

For a specified pair of values p_1 and p_2, we can find the sample sizes $n_1 = n_2 = n$ required to give the test of size α that has specified type II error β.

Approximate Sample Size for a Two-Sided Test on the Difference in Population Proportions

For the two-sided alternative, the common sample size is

$$n = \frac{\left[z_{\alpha/2}\sqrt{(p_1 + p_2)(q_1 + q_2)/2} + z_{\beta}\sqrt{p_1 q_1 + p_2 q_2}\right]^2}{(p_1 - p_2)^2} \tag{10.40}$$

where $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$.

For a one-sided alternative, replace $z_{\alpha/2}$ in Equation 10.40 by z_{α}.

10.6.3 Confidence Interval on the Difference in Population Proportions

The traditional confidence interval for $p_1 - p_2$ can be found directly because we know that

$$Z = \frac{\hat{P}_1 - \hat{P}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

is approximately a standard normal random variable. Thus $P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) \simeq 1 - \alpha$, so we can substitute for Z in this last expression and use an approach similar to the one employed previously to find an approximate $100(1-\alpha)\%$ two-sided confidence interval for $p_1 - p_2$.

Approximate Confidence Interval on the Difference in Population Proportions

If $\hat{p}_1$ and $\hat{p}_2$ are the sample proportions of observations in two independent random samples of sizes n_1 and n_2 that belong to a class of interest, an approximate two-sided $100(1-\alpha)\%$ confidence interval on the difference in the true proportions $p_1 - p_2$ is

$$\begin{aligned} &\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\ &\le p_1 - p_2 \le \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \end{aligned} \tag{10.41}$$

where $Z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution.

EXAMPLE 10.17 | Defective Bearings

Consider the process of manufacturing crankshaft bearings described in Example 8.7. Suppose that a modification is made in the surface finishing process and that, subsequently, a second random sample of 85 bearings is obtained. The number of defective bearings in this second sample is 8. Therefore, because $n_1 = 85$, $\hat{p}_1 = 10/85 = 0.1176, n_2 = 85$, and $\hat{p}_2 = 8/85 = 0.0941$, we can obtain an approximate 95% confidence interval on the difference in the proportion of defective bearings produced under the two processes from Equation 10.41 as follows:

$$\begin{aligned} &\hat{p}_1 - \hat{p}_2 - z_{0.025}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \le p_1 - p_2 \\ &\le \hat{p}_1 - \hat{p}_2 + z_{0.025}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \end{aligned}$$

or

$$\begin{aligned} &0.1176 - 0.0941 - 1.96\sqrt{\frac{0.1176(0.8824)}{85} + \frac{0.0941(0.9059)}{85}} \\ &\le p_1 - p_2 \le 0.1176 - 0.0941 \\ &+ 1.96\sqrt{\frac{0.1176(0.8824)}{85} + \frac{0.0941(0.9059)}{85}} \end{aligned}$$

This simplifies to

$$-0.0685 \le p_1 - p_2 \le 0.1155$$

Practical Interpretation: This confidence interval includes zero, so, based on the sample data, it seems unlikely that the changes made in the surface finish process have reduced the proportion of defective crankshaft bearings being produced.

The CI in Equation 10.41 is the **traditional** one usually given for a difference in two binomial proportions. However, the actual confidence level for this interval can deviate substantially from

the nominal or advertised value. So when we want a 95% CI (for example) and use $z_{0.025} = 1.96$ in Equation 10.41, the actual confidence level that we experience may differ from 95%. This situation can be improved by a very simple adjustment to the procedure: Add one success and one failure to the data from each sample and then calculate

$$\tilde{p}_1 = \frac{x_1 + 1}{n_1 + 2} \quad \text{and} \quad \tilde{n}_1 = n_1 + 2$$

$$\tilde{p}_2 = \frac{x_2 + 1}{n_2 + 2} \quad \text{and} \quad \tilde{n}_2 = n_2 + 2$$

Then replace $\hat{p}_1, \hat{p}_2, n_1,$ and n_2 by $\tilde{p}_1, \tilde{p}_2, \tilde{n}_1,$ and $\tilde{n}_2$ in Equation 10.41.

To illustrate how this works, reconsider the crankshaft bearing data from Example 10.17. Using the preceding procedure, we find that

$$\tilde{p}_1 = \frac{x_1 + 1}{n_1 + 2} = \frac{10 + 1}{85 + 2} = 0.1264 \quad \text{and} \quad \tilde{n}_1 = n_1 + 2 = 85 + 2 = 87$$

$$\tilde{p}_2 = \frac{x_2 + 1}{n_2 + 2} = \frac{8 + 1}{85 + 2} = 0.1034 \quad \text{and} \quad \tilde{n}_2 = n_2 + 2 = 85 + 2 = 87$$

If we now replace $\hat{p}_1, \hat{p}_2, n_1,$ and n_2 by $\tilde{p}_1, \tilde{p}_2, \tilde{n}_1,$ and $\tilde{n}_2$ in Equation 10.41, we find that the new improved CI is $-0.0730 \le p_1 - p_2 \le 0.1190$, which is similar to the traditional CI found in Example 10.17. The length of the traditional interval is 0.1840, and the length of the new and improved interval is 0.1920. The slightly longer interval is likely a reflection of the fact that the coverage of the improved interval is closer to the advertised level of 95%. However, because this CI also includes zero, the conclusions would be the same regardless of which CI is used.

10.7 Summary Table and Road Map for Inference Procedures for Two Samples

The table in the end pages of the book summarizes all of the two-sample parametric inference procedures given in this chapter. The table contains the null hypothesis statements, the test statistics, the criteria for rejection of the various alternative hypotheses, and the formulas for constructing the $100(1 - \alpha)\%$ confidence intervals.

The road map to select the appropriate parametric confidence interval formula or hypothesis test method for one-sample problems was presented in Table 8.1. In Table 10.5, we extend the road map to two-sample problems. The primary comments stated previously also apply here (except that we usually apply conclusions to a function of the parameters from each sample, such as the difference in means):

1. Determine the function of the parameters (and the distribution of the data) that is to be bounded by the confidence interval or tested by the hypothesis.
2. Check whether other parameters are known or need to be estimated (and whether any assumptions are made).

TABLE 10.5 Roadmap to Construct Confidence Intervals and Hypothesis Tests, Two-Sample Case

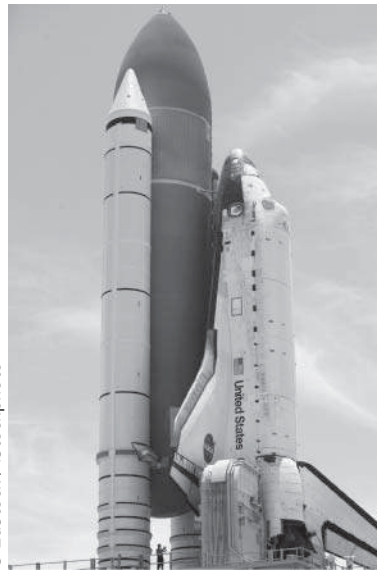
Function of the Parameters to be Bounded by the Confidence Interval or Tested with a Hypothesis	Symbol	Other Parameters?	Confidence Interval Section	Hypothesis Test Section	Comments
Difference in means from two normal distributions	$\mu_1 - \mu_2$	Standard deviations σ_1 and σ_2 known	10.1.3	10.1.1	
Difference in means from two arbitrary distributions with large sample sizes	$\mu_1 - \mu_2$	Sample sizes large enough that σ_1 and σ_2 are essentially known	10.1.3	10.1.1	Large sample size is often taken to be n_1 and $n_2 \geq 40$
Difference in means from two normal distributions	$\mu_1 - \mu_2$	Standard deviations σ_1 and σ_2 are unknown and assumed equal	10.2.3	10.2.1	Case 1: $\sigma_1 = \sigma_2$
Difference in means from two normal distributions	$\mu_1 - \mu_2$	Standard deviations σ_1 and σ_2 are unknown and NOT assumed equal	10.2.3	10.2.1	Case 2: $\sigma_1 \neq \sigma_2$
Difference in means from two normal distributions in a paired analysis	$\mu_D = \mu_1 - \mu_2$	Standard deviation of differences are unknown	10.4	10.4	Paired analysis calculates differences and uses a one-sample method for inference on the mean difference
Ratio of variances of two normal distributions	σ_1^2/σ_2^2	Means μ_1 and μ_2 unknown and estimated	10.5.4	10.5.2	
Difference in two population proportions	$p_1 - p_2$	None	10.6.3	10.6.1	Normal approximation to the binomial distribution used for the tests and confidence intervals

Important Terms and Concepts

Comparative experiments
Completely randomized experiment
Confidence intervals on differences and ratios
Critical region for a test statistic
Null and alternative hypothesis
Observational study
One-sided and two-sided alternative hypotheses
Operating characteristic curves
Paired *t*-test
Pooled *t*-test
P-value
Reference distribution for a test statistic
Sample size determination for hypothesis tests and confidence intervals
Statistical hypothesis
Test statistic
Treatments
Wilcoxon rank-sum test

CHAPTER **11**

Simple Linear Regression and Correlation



© EdStock / iStockphoto

LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Use simple linear regression for building empirical models to engineering and scientific data
2. Understand how the method of least squares is used to estimate the parameters in a linear regression model
3. Analyze residuals to determine whether the regression model is an adequate fit to the data or whether any underlying assumptions are violated
4. Test statistical hypotheses and construct confidence intervals on regression model parameters
5. Use the regression model to predict a future observation and to construct an appropriate prediction interval on the future observation
6. Apply the correlation model
7. Use simple transformations to achieve a linear regression model

CHAPTER OUTLINE

11.1 Empirical Models

11.2 Simple Linear Regression

11.3 Properties of the Least Squares Estimators

11.4 Hypothesis Tests in Simple Linear Regression

- **11.4.1** Use of *t*-Tests
- **11.4.2** Analysis of Variance Approach to Test Significance of Regression

11.5 Confidence Intervals

- **11.5.1** Confidence Intervals on the Slope and Intercept
- **11.5.2** Confidence Interval on the Mean Response

11.6 Prediction of New Observations

11.7 Adequacy of the Regression Model

- **11.7.1** Residual Analysis
- **11.7.2** Coefficient of Determination (R^2)

11.8 Correlation

11.9 Regression on Transformed Variables

11.10 Logistic Regression

The Space Shuttle *Challenger* accident in January 1986 was the result of the failure of O-rings used to seal field joints in the solid rocket motor because of the extremely low ambient temperatures at the time of launch. Prior to the launch, there were data on the occurrence of O-ring failure and the corresponding temperature on 24 prior launches or static firings of the motor. A statistical model relating the probability of O-ring failure to temperature would have provided a measure of the risk associated with launching at the low temperature at the time of the *Challenger* tragedy.

11.1 Empirical Models

Many problems in engineering and the sciences involve a study or analysis of the relationship between two or more variables. For example, the pressure of a gas in a container is related to the temperature, the velocity of water in an open channel is related to the width of the channel, and the displacement of a particle at a certain time is related to its velocity. In this last example, if we let d_0 be the displacement of the particle from the origin at time $t = 0$ and v be the velocity, the displacement at time t is $d_t = d_0 + vt$. This is an example of a **deterministic** linear relationship because (apart from measurement errors) the model predicts displacement perfectly.

However, in many situations, the relationship between variables is not deterministic. For example, the electrical energy consumption of a house (y) is related to the size of the house (x, in square feet), but it is unlikely to be a deterministic relationship. Similarly, the fuel usage of an automobile (y) is related to the vehicle weight x, but the relationship is not a deterministic one. In both of these examples, the value of the response of interest y (energy consumption, fuel usage) cannot be predicted perfectly from knowledge of the corresponding x. It is possible for different automobiles to have different fuel usage even if they weigh the same, and it is possible for different houses to use different amounts of electricity even if they are the same size.

The collection of statistical tools that are used to model and explore relationships between variables that are related in a nondeterministic manner is called **regression analysis**. Because problems of this type occur so frequently in many branches of engineering and science, regression analysis is one of the most widely used statistical tools. In this chapter, we present the situation in which there is only one independent or predictor variable x and the relationship with the response y is assumed to be linear. Although this seems to be a simple scenario, many practical problems fall into this framework.

For example, in a chemical process, suppose that the yield of the product is related to the process-operating temperature. Regression analysis can be used to build a model to predict yield at a given temperature level. This model can also be used for process optimization, such as finding the level of temperature that maximizes yield, or for process control purposes.

As an illustration, consider the data in Table 11.1. In this table, y is the purity of oxygen produced in a chemical distillation process, and x is the percentage of hydrocarbons present in the main condenser of the distillation unit. Figure 11.1 presents a **scatter diagram** of the data in Table 11.1. This is just a graph on which each (x_i, y_i) pair is represented as a point plotted in a two-dimensional coordinate system. This scatter diagram was produced by a computer, and we selected an option that shows dot diagrams of the x and y variables along the top and right margins of the graph, respectively, making it easy to see the distributions of the individual variables (box plots or histograms could also be selected). Inspection of this scatter diagram indicates that, although no simple curve will pass exactly through all the points, there is a strong indication that the points lie scattered randomly around a straight line. Therefore, it is probably reasonable to assume that the mean of the random variable Y is related to x by the following straight-line relationship:

$$E(Y|x) = \mu_{Y|x} = \beta_0 + \beta_1 x$$

where the slope and intercept of the line are called **regression coefficients**. Although the mean of Y is a linear function of x, the actual observed value y does not fall exactly on a straight line. The appropriate way to generalize this to a probabilistic linear model is to assume that the expected

TABLE 11.1 Oxygen and Hydrocarbon Levels

Observation Number	Hydrocarbon Level x(%)	Purity y(%)
1	0.99	90.01
2	1.02	89.05
3	1.15	91.43
4	1.29	93.74
5	1.46	96.73
6	1.36	94.45
7	0.87	87.59
8	1.23	91.77
9	1.55	99.42
10	1.40	93.65
11	1.19	93.54
12	1.15	92.52
13	0.98	90.56
14	1.01	89.54
15	1.11	89.85
16	1.20	90.39
17	1.26	93.25
18	1.32	93.41
19	1.43	94.98
20	0.95	87.33

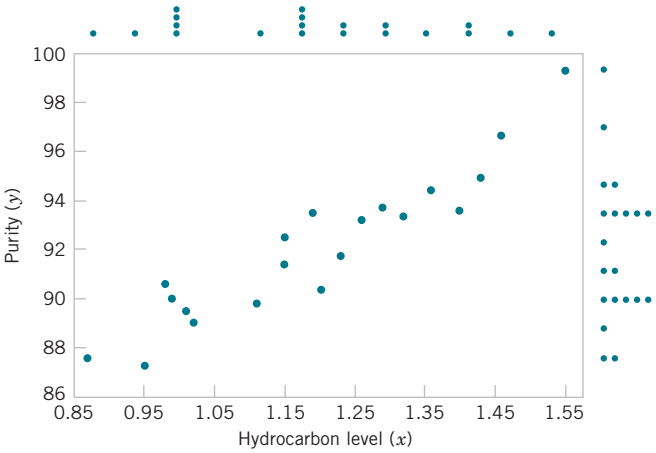


FIGURE 11.1
Scatter diagram of oxygen purity versus hydrocarbon level from Table 11.1.

value of Y is a linear function of x but that for a fixed value of x, the actual value of Y is determined by the mean value function (the linear model) plus a random error term, say,

$$Y = \beta_0 + \beta_1 x + \epsilon \tag{11.1}$$

where ϵ is the random error term. We call this model the **simple linear regression model** because it has only one independent variable or **regressor**. Sometimes a model like this arises from a theoretical relationship. At other times, we will have no theoretical knowledge of the relationship between x and y and will base the choice of the model on inspection of a scatter diagram, such as we did with the oxygen purity data. We then think of the regression model as an **empirical model**.

To gain more insight into this model, suppose that we can fix the value of x and observe the value of the random variable Y. Now if x is fixed, the random component ϵ on the right-hand side of the model in Equation 11.1 determines the properties of Y. Suppose that the mean and variance of ϵ are 0 and σ^2, respectively. Then,

$$E(Y|x) = E(\beta_0 + \beta_1 x + \epsilon) = \beta_0 + \beta_1 x + E(\epsilon) = \beta_0 + \beta_1 x$$

Notice that this is the same relationship that we initially wrote down empirically from inspection of the scatter diagram in Figure 11.1. The variance of Y given x is

$$V(Y|x) = V(\beta_0 + \beta_1 x + \epsilon) = V(\beta_0 + \beta_1 x) + V(\epsilon) = 0 + \sigma^2 = \sigma^2$$

Thus, the true regression model $\mu_{Y|x} = \beta_0 + \beta_1 x$ is a line of mean values; that is, the height of the regression line at any value of x is just the expected value of Y for that x. The slope, β_1, can be interpreted as the change in the mean of Y for a unit change in x. Furthermore, the variability of Y at a particular value of x is determined by the error variance σ^2. This implies that there is a distribution of Y-values at each x and that the variance of this distribution is the same at each x.

For example, suppose that the true regression model relating oxygen purity to hydrocarbon level is $\mu_{Y|x} = 75 + 15x$, and suppose that the variance is $\sigma^2 = 2$. Figure 11.2 illustrates this situation. Notice that we have used a normal distribution to describe the random variation in σ^2. Because σ^2 is the sum of a constant $\beta_0 + \beta_1 x$ (the mean) and a normally distributed random variable, Y is a normally distributed random variable. The variance σ^2 determines the variability in the observations Y on oxygen purity. Thus, when σ^2 is small, the observed values of Y will fall close to the line, and when σ^2 is large, the observed values of Y may deviate considerably from the line. Because σ^2 is constant, the variability in Y at any value of x is the same.

The regression model describes the relationship between oxygen purity Y and hydrocarbon level x. Thus, for any value of hydrocarbon level, oxygen purity has a normal distribution with mean $75 + 15x$ and variance 2. For example, if $x = 1.25$, Y has mean value $\mu_{Y|x} = 75 + 15(1.25) = 93.75$ and variance 2.

In most real-world problems, the values of the intercept and slope (β_0, β_1) and the error variance σ^2 will not be known and must be estimated from sample data. Then this fitted regression equation or model is typically used in prediction of future observations of Y, or for estimating the mean response at a particular level of x. To illustrate, a chemical engineer might be interested in estimating the mean purity of oxygen produced when the hydrocarbon level is $x = 1.25\%$. This chapter discusses such procedures and applications for the simple linear regression model. Chapter 12 discusses multiple linear regression models that involve more than one regressor.

Historical Note Sir Francis Galton first used the term **regression analysis** in a study of the heights of fathers (x) and sons (y). Galton fit a least squares line and used it to predict the son's height from the father's height. He found that if a father's height was above average, the son's height would also be above average but not by as much as the father's height was. A similar effect was observed for below-average heights. That is, the son's height "regressed" toward the average. Consequently, Galton referred to the least squares line as a **regression line**.

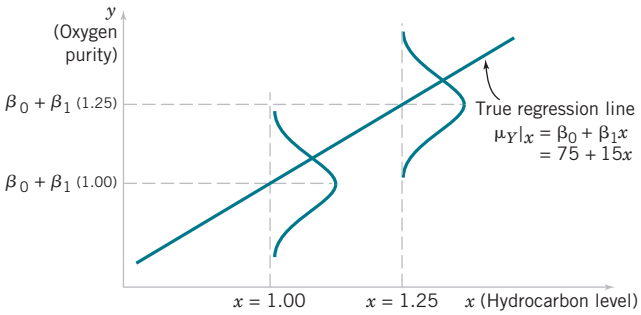


FIGURE 11.2
The distribution of Y for a given value of x for the oxygen purity-hydrocarbon data.

Abuses of Regression Regression is widely used and frequently misused; we mention several common abuses of regression briefly here. Care should be taken in selecting variables with which to construct regression equations and in determining the form of the model. It is possible to develop statistically significant relationships among variables that are completely unrelated in a **causal** sense. For example, we might attempt to relate the shear strength of spot welds with the number of empty parking spaces in the visitor parking lot. A straight line may even appear to provide a good fit to the data, but the relationship is an unreasonable one on which to rely. We cannot increase the weld strength by blocking off parking spaces. A strong observed association between variables does not necessarily imply that a causal relationship exists between them. This type of effect is encountered fairly often in retrospective data analysis and even in observational studies. *Designed experiments* are the only way to determine cause-and-effect relationships.

Regression relationships are valid for values of the regressor variable only within the range of the original data. The linear relationship that we have tentatively assumed may be valid over the original range of x, but it may be unlikely to remain so as we extrapolate—that is, if we use values of x beyond that range. In other words, as we move beyond the range of values of R^2 for which data were collected, we become less certain about the validity of the assumed model. Regression models are not necessarily valid for extrapolation purposes.

Now this does not mean *do not ever extrapolate*. For many problem situations in science and engineering, extrapolation of a regression model is the only way to even approach the problem. However, there is a strong warning to *be careful*. A modest extrapolation may be perfectly all right in many cases, but a large extrapolation will almost never produce acceptable results.

11.2 Simple Linear Regression

The case of simple linear regression considers a single regressor variable or predictor variable x and a dependent or **response variable** Y. Suppose that the true relationship between Y and x is a straight line and that the observation Y at each level of x is a random variable. As noted previously, the expected value of Y for each value of x is

$$E(Y|x) = \beta_0 + \beta_1 x$$

where the intercept β_0 and the slope β_1 are unknown regression coefficients. We assume that each observation, Y, can be described by the model

$$Y = \beta_0 + \beta_1 x + \epsilon \tag{11.2}$$

where ϵ is a random error with mean zero and (unknown) variance σ^2. The random errors corresponding to different observations are also assumed to be uncorrelated random variables.

Suppose that we have n pairs of observations (x_1, y_1), (x_2, y_2), … , (x_n, y_n). Figure 11.3 is a typical scatter plot of observed data and a candidate for the estimated regression line. The estimates of β_0 and β_1 should result in a line that is (in some sense) a "best fit" to the data. German scientist Karl Gauss (1777–1855) proposed estimating the parameters β_0 and β_1 in Equation 11.2 to minimize the sum of the squares of the vertical deviations in Figure 11.3.

We call this criterion for estimating the regression coefficients the method of **least squares**. Using Equation 11.2, we may express the n observations in the sample as

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \qquad i = 1, 2, \ldots, n \tag{11.3}$$

and the sum of the squares of the deviations of the observations from the true regression line is

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \tag{11.4}$$

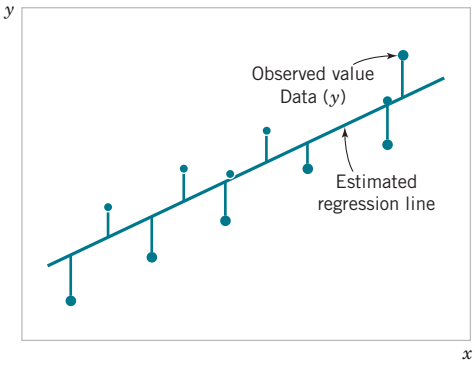


FIGURE 11.3
Deviations of the data from the estimated regression model.

The least squares estimators of β_0 and β_1, say, $\hat{\beta}_0$ and $\hat{\beta}_1$, must satisfy

$$\left.\frac{\partial L}{\partial \beta_0}\right|_{\hat{\beta}_0,\hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$
$$\left.\frac{\partial L}{\partial \beta_1}\right|_{\hat{\beta}_0,\hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)x_i = 0 \qquad (11.5)$$

Simplifying these two equations yields

$$n\hat{\beta}_0 + \hat{\beta}_1\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$
$$\hat{\beta}_0\sum_{i=1}^{n} x_i + \hat{\beta}_1\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i \qquad (11.6)$$

Equations 11.6 are called the **least squares normal equations**. The solution to the normal equations results in the least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$.

Least Squares Estimates

The **least squares estimates** of the intercept and slope in the simple linear regression model are

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \qquad (11.7)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \qquad (11.8)$$

where $\bar{y} = (1/n)\sum_{i=1}^{n} y_i$ and $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$.

The **fitted** or **estimated regression line** is therefore

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \qquad (11.9)$$

Note that each pair of observations satisfies the relationship

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i, \qquad i = 1, 2, \ldots, n$$

where $e_i = y_i - \hat{y}_i$ is called the **residual**. The residual describes the error in the fit of the model to the ith observation y_i. Later in this chapter, we use the residuals to provide information about the adequacy of the fitted model.

Notationally, it is occasionally convenient to give special symbols to the numerator and denominator of Equation 11.8. Given data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, let

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} \tag{11.10}$$

and

$$S_{xy} = \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n} \tag{11.11}$$

EXAMPLE 11.1 | Oxygen Purity

We now fit a simple linear regression model to the oxygen purity data in Table 11.1. The following quantities may be computed:

$$n = 20 \quad \sum_{i=1}^{20} x_i = 23.92 \quad \sum_{i=1}^{20} y_i = 1{,}843.21$$

$$\bar{x} = 1.1960 \quad \bar{y} = 92.1605$$

$$\sum_{i=1}^{20} y_i^2 = 170{,}044.5321 \quad \sum_{i=1}^{20} x_i^2 = 29.2892$$

$$\sum_{i=1}^{20} x_i y_i = 2{,}214.6566$$

$$S_{xx} = \sum_{i=1}^{20} x_i^2 - \frac{\left(\sum_{i=1}^{20} x_i\right)^2}{20} = 29.2892 - \frac{(23.92)^2}{20} = 0.68088$$

and

$$S_{xy} = \sum_{i=1}^{20} x_i y_i - \frac{\left(\sum_{i=1}^{20} x_i\right)\left(\sum_{i=1}^{20} y_i\right)}{20}$$

$$= 2{,}214.6566 - \frac{(23.92)(1{,}843.21)}{20} = 10.17744$$

Therefore, the least squares estimates of the slope and intercept are

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{10.17744}{0.68088} = 14.94748$$

and

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 92.1605 - (14.94748)1.196 = 74.28331$$

The fitted simple linear regression model (with the coefficients reported to three decimal places) is

$$\hat{y} = 74.283 + 14.947x$$

This model is plotted in Figure 11.4, along with the sample data.

Practical Interpretation: Using the regression model, we would predict oxygen purity of $\hat{y} = 89.23\%$ when the hydrocarbon level is $x = 1.00\%$. The 89.23% purity may be interpreted as an estimate of the true population mean purity when $x = 1.00\%$, or as an estimate of a new observation when $x = 1.00\%$. These estimates are, of course, subject to error; that is, it is unlikely that a future observation on purity would be exactly 89.23% when the hydrocarbon level is 1.00%. In subsequent sections, we see how to use confidence intervals and prediction intervals to describe the error in estimation from a regression model.

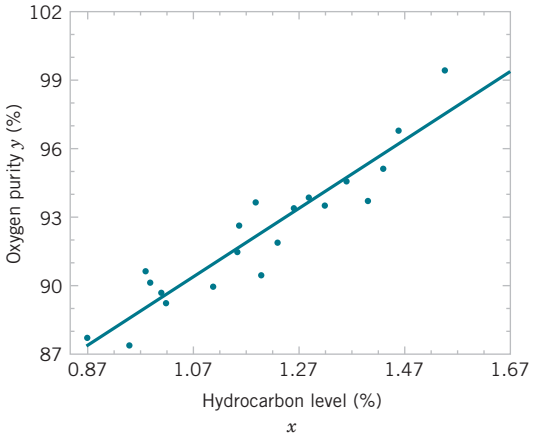


FIGURE 11.4

Scatter plot of oxygen purity y versus hydrocarbon level x and regression model $\hat{y} = 74.283 + 14.947x$.

TABLE 11.2 **Software Output for the Oxygen Purity Data in Example 11.1**

Purity = 74.3 + 14.9 HC Level

Predictor	Coef	SE Coef	T	P
Constant	74.283 ← $\hat{\beta}_0$	1.593	46.62	0.000
HC level	14.947 ← $\hat{\beta}_1$	1.317	11.35	0.000

S = 1.087 R-Sq = 87.7% R-Sq (adj) = 87.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	152.13	152.13	128.86	0.000
Residual error	18	21.25 ← SS_E	1.18 ← $\hat{\sigma}^2$		
Total	19	173.38			

Predicted Values for New Observations

New obs	Fit	SE Fit	95.0% CI	95.0% PI
1	89.231	0.354	(88.486, 89.975)	(86.830, 91.632)

Values of Predictors for New Observations

New obs	HC Level
1	1.00

Computer software programs are widely used in regression modeling. These programs typically carry more decimal places in the calculations. See Table 11.2 for a portion of typical output from a software package for this problem. The estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ are highlighted. In subsequent sections, we provide explanations for the information provided in this computer output.

Estimating σ^2 There is actually another unknown parameter in our regression model, σ^2 (the variance of the error term ϵ). The residuals $e_i = y_i - \hat{y}_i$ are used to obtain an estimate of σ^2. The sum of squares of the residuals, often called the **error sum of squares**, is

$$SS_E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{11.12}$$

We can show that the expected value of the error sum of squares is $E(SS_E) = (n-2)\sigma^2$. Therefore, an **unbiased estimator** of σ^2 is

Estimator of Variance

$$\hat{\sigma}^2 = \frac{SS_E}{n-2} \tag{11.13}$$

Computing SS_E using Equation 11.12 would be fairly tedious. A more convenient computing formula can be obtained by substituting $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ into Equation 11.12 and simplifying. The resulting computing formula is

$$SS_E = SS_T - \hat{\beta}_1 S_{xy} \tag{11.14}$$

where $SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$ is the total sum of squares of the response variable y. Formulas such as this are presented in Section 11.4. The error sum of squares and the estimate of σ^2 for the oxygen purity data, $\hat{\sigma}^2 = 1.18$, are highlighted in the computer output in Table 11.2.

11.3 Properties of the Least Squares Estimators

The statistical properties of the least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ may be easily described. Recall that we have assumed that the error term ϵ in the model $Y = \beta_0 + \beta_1 x + \epsilon$ is a random variable with mean zero and variance σ^2. Because the values of x are fixed, Y is a random variable with mean $\mu_{Y|x} = \beta_0 + \beta_1 x$ and variance σ^2. Therefore, the values of $\hat{\beta}_0$ and $\hat{\beta}_1$ depend on the observed y's; thus, the least squares estimators of the regression coefficients may be viewed as random variables. We now investigate the bias and variance properties of the least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$.

Consider first $\hat{\beta}_1$. Because $\hat{\beta}_1$ is a linear combination of the observations Y_i, we can use properties of expectation to show that the expected value of $\hat{\beta}_1$ is

$$E(\hat{\beta}_1) = \beta_1 \tag{11.15}$$

Thus, $\hat{\beta}_1$ is an **unbiased estimator** in simple linear regression of the true slope β_1.

Now consider the variance of $\hat{\beta}_1$. Because we have assumed that $V(\epsilon_i) = \sigma^2$, it follows that $V(Y_i) = \sigma^2$. Because $\hat{\beta}_1$ is a linear combination of the observations Y_i, the results in Section 5.7 can be applied to show that

$$V(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}} \tag{11.16}$$

For the intercept, we can show in a similar manner that

$$E(\hat{\beta}_0) = \beta_0 \quad \text{and} \quad V(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right] \tag{11.17}$$

Thus, $\hat{\beta}_0$ is an unbiased estimator of the intercept β_0. The covariance of the random variables $\hat{\beta}_0$ and $\hat{\beta}_1$ is not zero. It can be shown that $\text{cov}(\hat{\beta}_0, \hat{\beta}_1) = -\sigma^2 \bar{x} / S_{xx}$.

The estimate of σ^2 could be used in Equations 11.16 and 11.17 to provide estimates of the variance of the slope and the intercept. We call the square roots of the resulting variance estimators the **estimated standard errors** of the slope and intercept, respectively.

Estimated Standard Errors

In simple linear regression, the **estimated standard error of the slope** and the **estimated standard error of the intercept** are

$$se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \qquad \text{and} \qquad se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right]}$$

respectively, where $\hat{\sigma}^2$ is computed from Equation 11.13.

The computer output in Table 11.2 reports the estimated standard errors of the slope and intercept under the column heading SE coeff.

11.4 Hypothesis Tests in Simple Linear Regression

An important part of assessing the adequacy of a linear regression model is testing statistical hypotheses about the model parameters and constructing certain confidence intervals. Hypothesis testing in simple linear regression is discussed in this section, and Section 11.5 presents methods

for constructing confidence intervals. To test hypotheses about the slope and intercept of the regression model, we must make the additional assumption that the error component in the model, ϵ, is normally distributed. Thus, the complete assumptions are that the errors are normally and independently distributed with mean zero and variance σ^2, abbreviated NID $(0, \sigma^2)$.

11.4.1 Use of *t*-Tests

Suppose that we wish to test the hypothesis that the slope equals a constant, say, $\beta_{1,0}$. The appropriate hypotheses are

$$H_0: \beta_1 = \beta_{1,0} \qquad H_1: \beta_1 \neq \beta_{1,0} \tag{11.18}$$

where we have assumed a two-sided alternative. Because the errors ϵ_i are NID $(0, \sigma^2)$, it follows directly that the observations Y_i are NID$(\beta_0 + \beta_1 x_i, \sigma^2)$. Now $\hat{\beta}_1$ is a linear combination of independent normal random variables and, consequently, $\hat{\beta}_1$ is $N(\beta_1, \sigma^2/S_{xx})$, using the bias and variance properties of the slope discussed in Section 11.3. In addition, $(n-2)\hat{\sigma}^2/\sigma^2$ has a chi-square distribution with $n-2$ degrees of freedom, and $\hat{\beta}_1$ is independent of $\hat{\sigma}^2$. As a result of those properties, the statistic

Test Statistic for the Slope

$$T_0 = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2/S_{xx}}} \tag{11.19}$$

follows the t distribution with $n-2$ degrees of freedom under $H_0: \beta_1 = \beta_{1,0}$. We would reject $H_0: \beta_1 = \beta_{1,0}$ if

$$|t_0| > t_{\alpha/2,n-2} \tag{11.20}$$

where t_0 is computed from Equation 11.19. The denominator of Equation 11.19 is the standard error of the slope, so we could write the test statistic as

$$T_0 = \frac{\hat{\beta}_1 - \beta_{1,0}}{se(\hat{\beta}_1)}$$

A similar procedure can be used to test hypotheses about the intercept. To test

$$H_0: \beta_0 = \beta_{0,0} \qquad H_1: \beta_0 \neq \beta_{0,0} \tag{11.21}$$

we would use the statistic

$$T_0 = \frac{\hat{\beta}_0 - \beta_{0,0}}{\sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]}} = \frac{\hat{\beta}_0 - \beta_{0,0}}{se(\hat{\beta}_0)} \tag{11.22}$$

and reject the null hypothesis if the computed value of this test statistic, t_0, is such that $|t_0| > t_{\alpha/2,n-2}$. Note that the denominator of the test statistic in Equation 11.22 is just the standard error of the intercept.

A very important special case of the hypotheses of Equation 11.18 is

$$H_0: \beta_1 = 0 \qquad H_1: \beta_1 \neq 0 \tag{11.23}$$

These hypotheses relate to the **significance of regression**. Failure to reject $H_0: \beta_1 = 0$ is equivalent to concluding that there is no linear relationship between x and Y. This situation is illustrated in Figure 11.5. Note that this may imply either that x is of little value in explaining the variation in

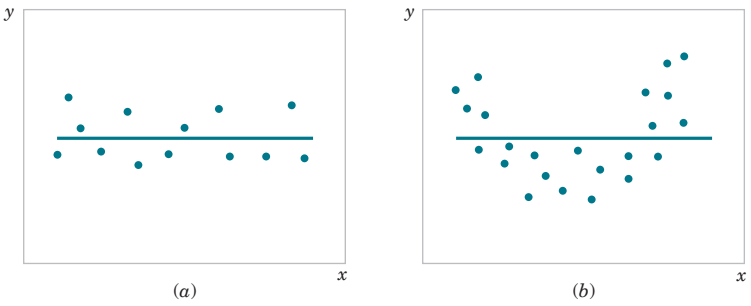


FIGURE 11.5

The hypothesis H_0: $\beta_1 = 0$ is not rejected.

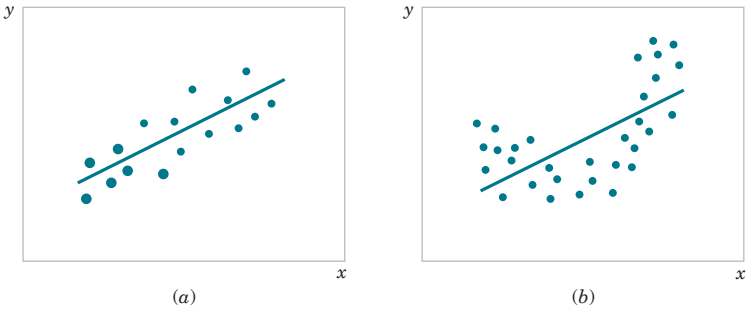


FIGURE 11.6

The hypothesis H_0: $\beta_1 = 0$ is rejected.

Y and that the best estimator of Y for any x is $\hat{y} = \overline{Y}$ [Figure 11.5(a)] or that the true relationship between x and Y is not linear [Figure 11.5(b)]. Alternatively, if H_0: $\beta_1 = 0$ is rejected, this implies that x is of value in explaining the variability in Y (see Figure 11.6). Rejecting H_0: $\beta_1 = 0$ could mean either that the straight-line model is adequate [Figure 11.6(a)] or that, although there is a linear effect of x, better results could be obtained with the addition of higher-order polynomial terms in x [Figure 11.6(b)].

EXAMPLE 11.2 | Oxygen Purity Tests of Coefficients

We now test for significance of regression using the model for the oxygen purity data from Example 11.1. The hypotheses are

$$H_0: \beta_1 = 0 \qquad H_1: \beta_1 \neq 0$$

and we use $\alpha = 0.01$. From Example 11.1 and Table 11.2 we have

$$\hat{\beta}_1 = 14.947, \qquad n = 20, \qquad S_{xx} = 0.68088, \qquad \hat{\sigma}^2 = 1.18$$

so the t-statistic in Equation 10.20 becomes

$$t_0 = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2/S_{xx}}} = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{14.947}{\sqrt{1.18/0.68088}} = 11.35$$

Practical Interpretation: Because the reference value of t is $t_{0.005,18} = 2.88$, the value of the test statistic is very far into the critical region, implying that H_0: $\beta_1 = 0$ should be rejected. There is strong evidence to support this claim. The P-value for this test is $P \simeq 1.23 \times 10^{-9}$. This was obtained manually with a calculator.

Table 11.2 presents the typical computer output for this problem. Notice that the t-statistic value for the slope is computed as 11.35 and that the reported P-value is $P = 0.000$. The computer also reports the t-statistic for testing the hypothesis H_0: $\beta_0 = 0$. This statistic is computed from Equation 11.22, with $\beta_{0,0} = 0$, as $t_0 = 46.62$. Clearly, then, the hypothesis that the intercept is zero is rejected.

11.4.2 Analysis of Variance Approach to Test Significance of Regression

A method called the **analysis of variance** can be used to test for significance of regression. The procedure partitions the total variability in the response variable into meaningful components as the basis for the test. The **analysis of variance identity** is as follows:

Analysis of Variance Identity

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \tag{11.24}$$

The two components on the right-hand side of Equation 11.24 measure, respectively, the amount of variability in y_i accounted for by the regression line and the residual variation left unexplained by the regression line. We usually call $SS_E=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$ the **error sum of squares** and $SS_R=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$ the **regression sum of squares**. Symbolically, Equation 11.24 may be written as

$$SS_T=SS_R+SS_E \tag{11.25}$$

where $SS_T=\sum_{i=1}^{n}(y_i-\bar{y})^2$ is the **total corrected sum of squares** of y. In Section 11.2, we noted that $SS_E=SS_T-\hat{\beta}_1S_{xy}$ (see Equation 11.14), so because $SS_T=\hat{\beta}_1S_{xy}+SS_E$, we note that the regression sum of squares in Equation 11.25 is $SS_R=\hat{\beta}_1S_{xy}$. The total sum of squares SS_T has $n-1$ degrees of freedom, and SS_R and SS_E have 1 and $n-2$ degrees of freedom, respectively.

We may show that $E[SS_E/(n-2)]=\sigma^2$ and $E(SS_R)=\sigma^2+\beta_1^2S_{xx}$ and that SS_E/σ^2 and SS_R/σ^2 are independent chi-square random variables with $n-2$ and 1 degrees of freedom, respectively. Thus, if the null hypothesis H_0: $\beta_1=0$ is true, the statistic

Test for Significance of Regression

$$F_0=\frac{SS_R/1}{SS_E/(n-2)}=\frac{MS_R}{MS_E} \tag{11.26}$$

follows the $F_{1,n-2}$ distribution, and we would reject H_0 if $f_0>f_{\alpha,1,n-2}$. The quantities $MS_R=SS_R/1$ and $MS_E=SS_E/(n-2)$ are called **mean squares**. In general, a mean square is always computed by dividing a sum of squares by its number of degrees of freedom. The test procedure is usually arranged in an **analysis of variance table**, such as Table 11.3.

TABLE 11.3 Analysis of Variance for Testing Significance of Regression

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Regression	$SS_R=\hat{\beta}_1S_{xy}$	1	MS_R	MS_R/MS_E
Error	$SS_E=SS_T-\hat{\beta}_1S_{xy}$	$n-2$	MS_E	
Total	SS_T	$n-1$		

Note that $MS_E=\hat{\sigma}^2$.

EXAMPLE 11.3 | Oxygen Purity ANOVA

We now use the analysis of variance approach to test for significance of regression using the oxygen purity data model from Example 11.1. Recall that $SS_T = 173.38$, $\hat{\beta}_1 = 14.947$, $S_{xy} = 10.17744$, and $n = 20$. The regression sum of squares is

$$SS_R = \hat{\beta}_1 S_{xy} = (14.947)10.17744 = 152.13$$

and the error sum of squares is

$$SS_E = SS_T - SS_R = 173.38 - 152.13 = 21.25$$

The analysis of variance for testing H_0: $\beta_1 = 0$ is summarized in the computer output in Table 11.2. The test statistic is $f_0 = MS_R/MS_E = 152.13/1.18 = 128.86$, for which we find that the P-value is $P \simeq 1.23 \times 10^{-9}$, so we conclude that β_1 is not zero.

Frequently computer packages have minor differences in terminology. For example, sometimes the regression sum of squares is called the "model" sum of squares and the error sum of squares is called the "residual" sum of squares.

Note that the analysis of variance procedure for testing for significance of regression is equivalent to the t-test in Section 11.4.1. That is, either procedure will lead to the same conclusions. This is easy to demonstrate by starting with the t-test statistic in Equation 11.19 with $\beta_{1,0} = 0$, say

$$T_0 = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2/S_{xx}}} \tag{11.27}$$

Squaring both sides of Equation 11.27 and using the fact that $\hat{\sigma}^2 = MS_E$ results in

$$T_0^2 = \frac{\hat{\beta}_1^2 S_{xx}}{MS_E} = \frac{\hat{\beta}_1 S_{xy}}{MS_E} = \frac{MS_R}{MS_E} \tag{11.28}$$

Note that T_0^2 in Equation 11.28 is identical to F_0 in Equation 11.26. It is true, in general, that the square of a t random variable with v degrees of freedom is an F random variable with 1 and v degrees of freedom in the numerator and denominator, respectively. Thus, the test using T_0 is equivalent to the test based on F_0. Note, however, that the t-test is somewhat more flexible in that it would allow testing against a one-sided and two-sided alternative hypothesis, while the F-test is restricted to a two-sided alternative.

11.5 Confidence Intervals

11.5.1 Confidence Intervals on the Slope and Intercept

In addition to point estimates of the slope and intercept, it is possible to obtain **confidence interval** estimates of these parameters. The width of these confidence intervals is a measure of the overall quality of the regression line. If the error terms, ϵ_i, in the regression model are normally and independently distributed,

$$(\hat{\beta}_1 - \beta_1)\Big/\sqrt{\hat{\sigma}^2/S_{xx}} \quad \text{and} \quad (\hat{\beta}_0 - \beta_0)\Big/\sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]}$$

are both distributed as t random variables with $n - 2$ degrees of freedom. This leads to the following definition of $100(1 - \alpha)\%$ confidence intervals on the slope and intercept.

Confidence Intervals on Parameters

Under the assumption that the observations are normally and independently distributed, a $100(1-\alpha)\%$ **confidence interval on the slope** β_1 in simple linear regression is

$$\hat{\beta}_1 - t_{\alpha/2,n-2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \le \beta_1 \le \hat{\beta}_1 + t_{\alpha/2,n-2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \tag{11.29}$$

Similarly, a $100(1-\alpha)\%$ **confidence interval on the intercept** β_0 is

$$\hat{\beta}_0 - t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right]} \le \beta_0 \le \hat{\beta}_0 + t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right]} \tag{11.30}$$

EXAMPLE 11.4 | Oxygen Purity Confidence Interval on the Slope

We find a 95% confidence interval on the slope of the regression line using the data in Example 11.1. Recall that $\hat{\beta}_1 = 14.947$, $S_{xx} = 0.68088$, and $\hat{\sigma}^2 = 1.18$ (see Table 11.2). Then, from Equation 11.29, we find

$$\hat{\beta}_1 - t_{0.025,18}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \le \beta_1 \le \hat{\beta}_1 + t_{0.025,18}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$$

or

$$14.947 - 2.101\sqrt{\frac{1.18}{0.68088}} \le \beta_1 \le 14.947 + 2.101\sqrt{\frac{1.18}{0.68088}}$$

This simplifies to

$$12.181 \le \beta_1 \le 17.713$$

Practical Interpretation: This CI does not include zero, so there is strong evidence (at $\alpha = 0.05$) that the slope is not zero. The CI is reasonably narrow (± 2.766) because the error variance is fairly small.

11.5.2 Confidence Interval on the Mean Response

A confidence interval may be constructed on the mean response at a specified value of x, say, x_0. This is a confidence interval about $E(Y|x_0) = \mu_{Y|x_0}$ and is sometimes referred to as a confidence interval about the regression line. Because $E(Y|x_0) = \mu_{Y|x_0} = \beta_0 + \beta_1 x_0$, we may obtain a point estimate of the mean of Y at $x = x_0$ ($\mu_{Y|x_0}$) from the fitted model as

$$\hat{\mu}_{Y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

Now $\hat{\mu}_{Y|x_0}$ is an unbiased point estimator of $\mu_{Y|x_0}$ because $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1. The variance of $\hat{\mu}_{Y|x_0}$ is

$$V(\hat{\mu}_{Y|x_0}) = \sigma^2\left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]$$

This last result follows from the fact that $\hat{\mu}_{Y|x_0} = \bar{y} + \hat{\beta}_1(x_0 - \bar{x})$ and $\text{cov}(\bar{Y}, \hat{\beta}_1) = 0$. Also, $\hat{\mu}_{Y|x_0}$ is normally distributed because $\hat{\beta}_1$ and $\hat{\beta}_0$ are normally distributed, and if we use $\hat{\sigma}^2$ as an estimate of σ^2, it is easy to show that

$$\frac{\hat{\mu}_{Y|x_0} - \mu_{Y|x_0}}{\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]}}$$

has a t distribution with $n-2$ degrees of freedom. This leads to the following confidence interval definition.

Confidence Interval on the Mean Response

A $100(1-\alpha)\%$ **confidence interval on the mean response** at the value of $x = x_0$, say $\mu_{Y|x_0}$, is given by

$$\hat{\mu}_{Y|x_0} - t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]} \le \mu_{Y|x_0} \le \hat{\mu}_{Y|x_0} + t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]} \tag{11.31}$$

where $\hat{\mu}_{Y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$ is computed from the fitted regression model.

Note that the width of the CI for $\mu_{Y|x_0}$ is a function of the value specified for x_0. The interval width is a minimum for $x_0 = \bar{x}$ and widens as $|x_0 - \bar{x}|$ increases.

EXAMPLE 11.5 | Oxygen Purity Confidence Interval on the Mean Response

We construct a 95% confidence interval about the mean response for the data in Example 11.1. The fitted model is $\hat{\mu}_{Y|x_0} = 74.283 + 14.947x_0$, and the 95% confidence interval on $\mu_{Y|x_0}$ is found from Equation 11.31 as

$$\hat{\mu}_{Y|x_0} \pm 2.101\sqrt{1.18\left[\frac{1}{20}+\frac{(x_0-1.1960)^2}{0.68088}\right]}$$

Suppose that we are interested in predicting mean oxygen purity when $x_0 = 1.00\%$. Then

$$\hat{\mu}_{Y|x_{1.00}} = 74.283 + 14.947(1.00) = 89.23$$

and the 95% confidence interval is

$$89.23 \pm 2.101\sqrt{1.18\left[\frac{1}{20}+\frac{(1.00-1.1960)^2}{0.68088}\right]}$$

or

$$89.23 \pm 0.75$$

Therefore, the 95% CI on $\mu_{Y|1.00}$ is

$$88.48 \le \mu_{Y|1.00} \le 89.98$$

This is a reasonably narrow CI.

Most computer software will also perform these calculations. Refer to Table 11.2. The predicted value of y at $x = 1.00$ is shown along with the 95% CI on the mean of y at this level of x.

By repeating these calculations for several different values for x_0, we can obtain confidence limits for each corresponding value of $\mu_{Y|x_0}$. Figure 11.7 is a display of the scatter diagram with the fitted model and the corresponding 95% confidence limits plotted as the upper and lower lines. The 95% confidence level applies only to the interval obtained at one value of x, not to the entire set of x-levels. Notice that the width of the confidence interval on $\mu_{Y|x_0}$ increases as $|x_0 - \bar{x}|$ increases.

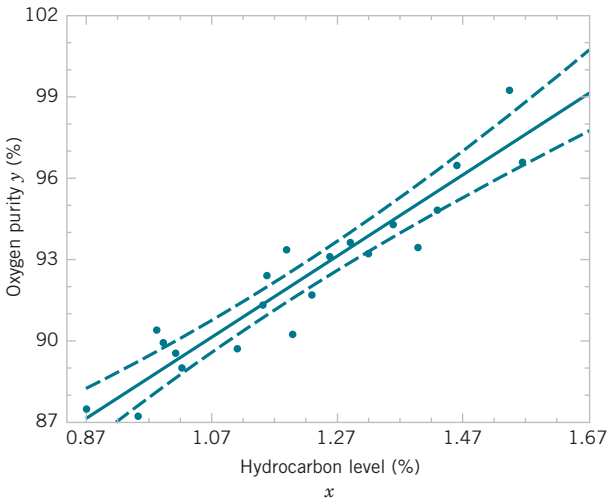


FIGURE 11.7

Scatter diagram of oxygen purity data from Example 11.1 with fitted regression line and 95% confidence limits on $\mu_{Y|x_0}$.

11.6 Prediction of New Observations

An important application of a regression model is predicting new or future observations Y corresponding to a specified level of the regressor variable x. If x_0 is the value of the regressor variable of interest,

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 \tag{11.32}$$

is the **point estimator** of the new or future value of the response Y_0.

Now consider obtaining an interval estimate for this future observation Y_0. This new observation is independent of the observations used to develop the regression model. Therefore, the confidence interval for $\mu_{Y|x_0}$ in Equation 11.31 is inappropriate because it is based only on the data used to fit the regression model. The confidence interval about $\mu_{Y|x_0}$ refers to the true mean response at $x = x_0$ (that is, a population parameter), not to future observations.

Let Y_0 be the future observation at $x = x_0$, and let $\hat{Y}_0$ given by Equation 11.32 be the estimator of Y_0. Note that the error in prediction

$$e_{\hat{p}} = Y_0 - \hat{Y}_0$$

is a normally distributed random variable with mean zero and variance

$$V(e_{\hat{p}}) = V(Y_0 - \hat{Y}_0) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]$$

because Y_0 is independent of $\hat{Y}_0$. If we use $\hat{\sigma}^2$ to estimate σ^2, we can show that

$$\frac{Y_0 - \hat{Y}_0}{\sqrt{\hat{\sigma}^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]}}$$

has a t distribution with $n - 2$ degrees of freedom. From this, we can develop the following **prediction interval** definition.

Prediction Interval

A $100(1 - \alpha)\%$ **prediction interval on a future observation** Y_0 at the value x_0 is given by

$$\hat{y}_0 - t_{\alpha/2,n-2} \sqrt{\hat{\sigma}^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]} \le Y_0 \le \hat{y}_0 + t_{\alpha/2,n-2} \sqrt{\hat{\sigma}^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]} \tag{11.33}$$

The value $\hat{y}_0$ is computed from the regression model $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$.

Notice that the prediction interval is of minimum width at $x_0 = \bar{x}$ and widens as $|x_0 - \bar{x}|$ increases. By comparing Equation 11.33 with Equation 11.31, we observe that the prediction interval at the point x_0 is always wider than the confidence interval at x_0. This results because the prediction interval depends on both the error from the fitted model and the error associated with future observations.

EXAMPLE 11.6 | Oxygen Purity Prediction Interval

To illustrate the construction of a prediction interval, suppose that we use the data in Example 11.1 and find a 95% prediction interval on the next observation of oxygen purity at $x_0 = 1.00\%$. Using Equation 11.33 and recalling from Example 11.5 that $\hat{y}_0 = 89.23$, we find that the prediction interval is

$$89.23 - 2.101\sqrt{1.18\left[1 + \frac{1}{20} + \frac{(1.00 - 1.1960)^2}{0.68088}\right]}$$
$$\leq Y_0 \leq 89.23 + 2.101\sqrt{1.18\left[1 + \frac{1}{20} + \frac{(1.00 - 1.1960)^2}{0.68088}\right]}$$

which simplifies to

$$86.83 \leq Y_0 \leq 91.63$$

This is a reasonably narrow prediction interval.

Typical computer software will also calculate prediction intervals. Refer to the output in Table 11.2. The 95% PI on the future observation at $x_0 = 1.00$ is shown in the display.

By repeating the foregoing calculations at different levels of x_0, we may obtain the 95% prediction intervals shown graphically as the lower and upper lines about the fitted regression model in Figure 11.8. Notice that this graph also shows the 95% confidence limits on $\mu_{Y|x_0}$ calculated in Example 11.5. It illustrates that the prediction limits are always wider than the confidence limits.

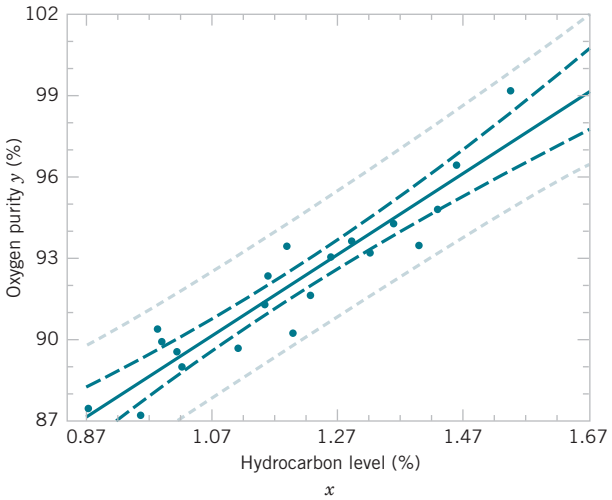


FIGURE 11.8

Scatter diagram of oxygen purity data from Example 11.1 with fitted regression line, 95% prediction limits (outer lines), and 95% confidence limits on $\mu_{Y|x_0}$.

11.7 Adequacy of the Regression Model

Fitting a regression model requires making several **assumptions**. Estimating the model parameters requires assuming that the errors are uncorrelated random variables with mean zero and constant variance. Tests of hypotheses and interval estimation require that the errors be normally distributed. In addition, we assume that the order of the model is correct; that is, if we fit a simple linear regression model, we are assuming that the phenomenon actually behaves in a linear or first-order manner.

The analyst should always consider the validity of these assumptions to be doubtful and conduct analyses to examine the adequacy of the model that has been tentatively entertained. In this section, we discuss methods useful in this respect.

11.7.1 Residual Analysis

The **residuals** from a regression model are $e_i = y_i - \hat{y}_i, i = 1, 2, \ldots, n$ where y_i is an actual observation and $\hat{y}_i$ is the corresponding fitted value from the regression model. Analysis of the residuals is frequently helpful in checking the assumption that the errors are approximately normally distributed with constant variance and in determining whether additional terms in the model would be useful.

As an approximate check of normality, the experimenter can construct a frequency histogram of the residuals or a **normal probability plot of residuals**. Many computer programs will produce a normal probability plot of residuals, and because the sample sizes in regression are often too small for a histogram to be meaningful, the normal probability plotting method is preferred.

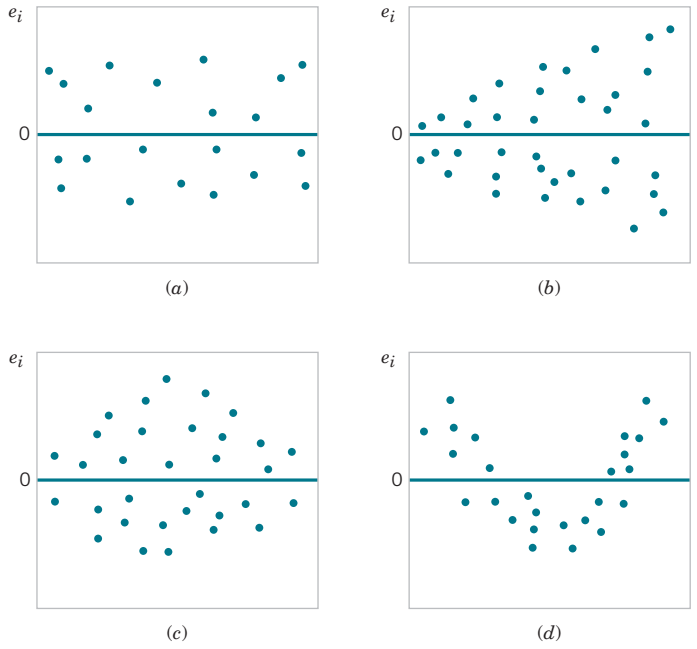


FIGURE 11.9

Patterns for residual plots. (a) Satisfactory. (b) Funnel. (c) Double bow. (d) Nonlinear.

Adapted from Montgomery, Peck, and Vining (2012).

It requires judgment to assess the abnormality of such plots. (Refer to the discussion of the "fat pencil" method in Section 6.7.)

We may also **standardize** the residuals by computing $d_i = e_i/\sqrt{\hat{\sigma}^2}, i = 1, 2, \ldots, n$. If the errors are normally distributed, approximately 95% of the standardized residuals should fall in the interval $(-2, +2)$. Residuals that are far outside this interval may indicate the presence of an **outlier**, that is, an observation that is not typical of the rest of the data. Various rules have been proposed for discarding outliers. However, they sometimes provide important information about unusual circumstances of interest to experimenters and should not be automatically discarded. For further discussion of outliers, see Montgomery, Peck, and Vining (2012).

It is frequently helpful to plot the residuals (1) in time sequence (if known), (2) against the $\hat{y}_i$, and (3) against the independent variable x. These graphs will usually look like one of the four general patterns shown in Figure 11.9. Pattern (a) in Figure 11.9 represents the ideal situation, and patterns (b), (c), and (d) represent anomalies. If the residuals appear as in (b), the variance of the observations may be increasing with time or with the magnitude of y_i or x_i. Data transformation on the response y is often used to eliminate this problem. Widely used variance-stabilizing transformations include the use of $\sqrt{y}$, $\ln y$, or $1/y$ as the response. See Montgomery, Peck, and Vining (2012) for more details regarding methods for selecting an appropriate transformation. Plots of residuals against $\hat{y}_i$ and x_i that look like (c) also indicate inequality of variance. Residual plots that look like (d) indicate model inadequacy; that is, higher-order terms should be added to the model, a transformation on the x-variable or the y-variable (or both) should be considered, or other regressors should be considered.

EXAMPLE 11.7 | Oxygen Purity Residuals

The regression model for the oxygen purity data in Example 11.1 is $\hat{y} = 74.283 + 14.947x$. Table 11.4 presents the observed and predicted values of y at each value of x from this data set along with the corresponding residual. These values were calculated using a computer and show the number of decimal places typical of computer output.

A normal probability plot of the residuals is shown in Figure 11.10. Because the residuals fall approximately along a straight line in the figure, we conclude that there is no severe departure from normality. The residuals are also plotted against the predicted value $\hat{y}_i$ in Figure 11.11 and against the hydrocarbon levels x_i in Figure 11.12. These plots do not indicate any serious model inadequacies.

TABLE 11.4 Oxygen Purity Data from Example 11.1, Predicted $\hat{y}$ Values, and Residuals

	Hydrocarbon Level, x	Oxygen Purity, y	Predicted Value, $\hat{y}$	Residual $e = y - \hat{y}$		Hydrocarbon Level, x	Oxygen Purity, y	Predicted Value, $\hat{y}$	Residual $e = y - \hat{y}$
1	0.99	90.01	89.081	0.929	11	1.19	93.54	92.071	1.469
2	1.02	89.05	89.530	−0.480	12	1.15	92.52	91.473	1.047
3	1.15	91.43	91.473	−0.043	13	0.98	90.56	88.932	1.628
4	1.29	93.74	93.566	0.174	14	1.01	89.54	89.380	0.160
5	1.46	96.73	96.107	0.623	15	1.11	89.85	90.875	−1.025
6	1.36	94.45	94.612	−0.162	16	1.20	90.39	92.220	−1.830
7	0.87	87.59	87.288	0.302	17	1.26	93.25	93.117	0.133
8	1.23	91.77	92.669	−0.899	18	1.32	93.41	94.014	−0.604
9	1.55	99.42	97.452	1.968	19	1.43	94.98	95.658	−0.678
10	1.40	93.65	95.210	−1.560	20	0.95	87.33	88.483	−1.153

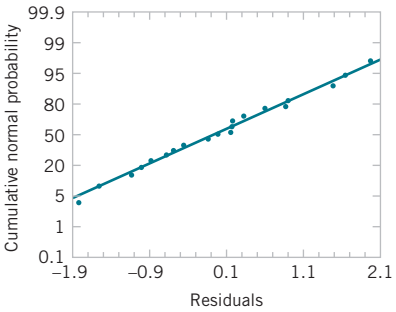


FIGURE 11.10
Normal probability plot of residuals.

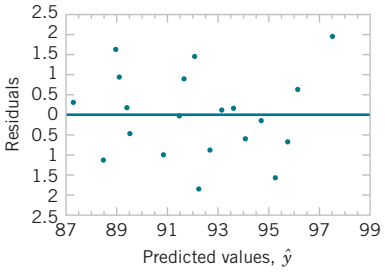


FIGURE 11.11
Plot of residuals versus predicted oxygen purity $\hat{y}$.

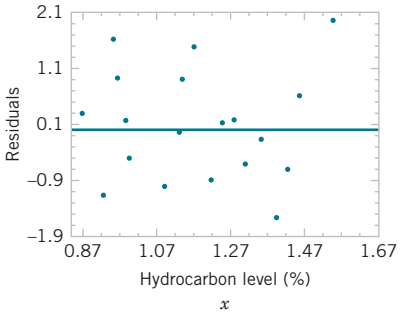


FIGURE 11.12
Plot of residuals versus hydrocarbon level x.

11.7.2 Coefficient of Determination (R^2)

A widely used measure for a regression model is the following ratio of sum of squares.

R^2

The **coefficient of determination** is

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{11.34}$$

The coefficient is often used to judge the adequacy of a regression model. Subsequently, we see that in the case in which X and Y are jointly distributed random variables, R^2 is the square of the correlation coefficient between X and Y. From the analysis of variance identity in Equations 11.24 and 11.25, $0 \le R^2 \le 1$. We often refer loosely to R^2 as the amount of variability in the data explained or accounted for by the regression model. For the oxygen purity regression model,

we have $R^2 = SS_R/SS_T = 152.13/173.38 = 0.877$; that is, the model accounts for 87.7% of the variability in the data.

The statistic R^2 should be used with caution because it is always possible to make R^2 unity by simply adding enough terms to the model. For example, we can obtain a "perfect" fit to n data points with a polynomial of degree $n - 1$. In general, R^2 will increase if we add a variable to the model, but this does not necessarily imply that the new model is superior to the old one. Unless the error sum of squares in the new model is reduced by an amount equal to the original error mean square, the new model will have a larger error mean square than the old one because of the loss of 1 error degree of freedom. Thus, the new model will actually be worse than the old one. The magnitude of R^2 is also impacted by the dispersion of the variable x. The larger the dispersion, the larger the value of R^2 will usually be.

There are several misconceptions about R^2. In general, R^2 does not measure the magnitude of the slope of the regression line. A large value of R^2 does not imply a steep slope. Furthermore, R^2 does not measure the appropriateness of the model because it can be artificially inflated by adding higher-order polynomial terms in x to the model. Even if y and x are related in a nonlinear fashion, R^2 will often be large. For example, R^2 for the regression equation in Figure 11.6(b) will be relatively large even though the linear approximation is poor. Finally, even though R^2 is large, this does not necessarily imply that the regression model will provide accurate predictions of future observations.

11.8 Correlation

Our development of regression analysis has assumed that x is a mathematical variable, measured with negligible error, and that Y is a random variable. Many applications of regression analysis involve situations in which both X and Y are random variables. In these situations, it is usually assumed that the observations (X_i, Y_i), $i = 1, 2, \ldots, n$ are jointly distributed random variables obtained from the distribution $f(x, y)$.

For example, suppose that we wish to develop a regression model relating the shear strength of spot welds to the weld diameter. In this example, we cannot control weld diameter. We would randomly select n spot welds and observe a diameter (X_i) and a shear strength (Y_i) for each. Therefore (X_i, Y_i) are jointly distributed random variables.

We assume that the joint distribution of X_i and Y_i is the bivariate normal distribution presented in Chapter 5, and μ_Y and σ_Y^2 are the mean and variance of Y, μ_X, σ_X^2 are the mean and variance of X, and ρ is the **correlation coefficient** between Y and X. Recall that the correlation coefficient is defined as

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{11.35}$$

where σ_{XY} is the covariance between Y and X.

The conditional distribution of Y for a given value of $X = x$ is

$$f_{Y|x}(y) = \frac{1}{\sqrt{2\pi}\sigma_{Y|x}} \exp\left[-\frac{1}{2}\left(\frac{y - \beta_0 - \beta_1 x}{\sigma_{Y|x}}\right)^2\right] \tag{11.36}$$

where

$$\beta_0 = \mu_Y - \mu_X \rho \frac{\sigma_Y}{\sigma_X} \tag{11.37}$$

$$\beta_1 = \frac{\sigma_Y}{\sigma_X} \rho \tag{11.38}$$

and the variance of the conditional distribution of Y given $X = x$ is

$$\sigma_{Y|x}^2 = \sigma_Y^2(1 - \rho^2) \tag{11.39}$$

That is, the conditional distribution of Y given $X = x$ is normal with mean

$$E(Y|x) = \beta_0 + \beta_1 x \tag{11.40}$$

and variance $\sigma^2_{Y|x}$. Thus, the mean of the conditional distribution of Y given $X = x$ is a simple linear regression model. Furthermore, a relationship exists between the correlation coefficient ρ and the slope β_1. From Equation 11.38, we see that if $\rho = 0$, then $\beta_1 = 0$, which implies that there is no regression of Y on X. That is, knowledge of X does not assist us in predicting Y.

The method of maximum likelihood may be used to estimate the parameters β_0 and β_1. It can be shown that the maximum likelihood estimators of those parameters are

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} \tag{11.41}$$

and

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} Y_i(X_i - \overline{X})}{\sum_{i=1}^{n} (X_i - \overline{X})^2} = \frac{S_{XY}}{S_{XX}} \tag{11.42}$$

We note that the estimators of the intercept and slope in Equations 11.41 and 11.42 are identical to those given by the method of least squares in the case in which X was assumed to be a mathematical variable. That is, the regression model with Y and X jointly normally distributed is equivalent to the model with X considered as a mathematical variable. This follows because the random variables Y given $X = x$ are independently and normally distributed with mean $\beta_0 + \beta_1 x$ and constant variance $\sigma^2_{Y|x}$. These results will also hold for any joint distribution of Y and X such that the conditional distribution of Y given X is normal.

It is possible to draw inferences about the correlation coefficient ρ in this model. The estimator of ρ is the **sample correlation coefficient**

$$\hat{\rho} = \frac{\sum_{i=1}^{n} Y_i(X_i - \overline{X})}{\left[\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2\right]^{1/2}} = \frac{S_{XY}}{\left(S_{XX} SS_T\right)^{1/2}} \tag{11.43}$$

Note that

$$\hat{\beta}_1 = \left(\frac{SS_T}{S_{XX}}\right)^{1/2} \hat{\rho} \tag{11.44}$$

so the slope $\hat{\beta}_1$ is just the sample correlation coefficient $\hat{\rho}$ multiplied by a scale factor that is the square root of the "spread" of the Y values divided by the "spread" of the X values. Thus, $\hat{\beta}_1$ and $\hat{\rho}$ are closely related, although they provide somewhat different information. The sample correlation coefficient $\hat{\rho}$ measures the linear association between Y and X, and $\hat{\beta}_1$ measures the predicted change in the mean of Y for a unit change in X. In the case of a mathematical variable x, $\hat{\rho}$ has no meaning because the magnitude of $\hat{\rho}$ depends on the choice of spacing of x. We may also write, from Equation 11.44,

$$\hat{\rho}^2 = \hat{\beta}_1^2 \frac{S_{XX}}{SS_T} = \frac{\hat{\beta}_1 S_{XY}}{SS_T} = \frac{SS_R}{SS_T}$$

which is just the coefficient of determination. That is, the coefficient of determination R^2 is just the square of the sample correlation coefficient between Y and X.

It is often useful to test the hypotheses

$$H_0\colon \rho = 0 \qquad H_1\colon \rho \neq 0 \tag{11.45}$$

The appropriate test statistic for these hypotheses is

Test Statistic for Zero Correlation

$$T_0 = \frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}} \tag{11.46}$$

which has the t distribution with $n-2$ degrees of freedom if H_0: $\rho = 0$ is true. Therefore, we would reject the null hypothesis if $|t_0| > t_{\alpha/2,n-2}$. This test is equivalent to the test of the hypothesis H_0: $\beta_1 = 0$ given in Section 11.5.1. This equivalence follows directly from Equation 11.46.

The test procedure for the hypotheses

$$H_0\text{: } \rho = \rho_0 \qquad H_1\text{: } \rho \neq \rho_0 \tag{11.47}$$

where $\rho_0 \neq 0$ is somewhat more complicated. For moderately large samples (say, $n \geq 25$), the statistic

$$Z = \operatorname{arctanh} \hat{\rho} = \frac{1}{2} \ln \frac{1+\hat{\rho}}{1-\hat{\rho}} \tag{11.48}$$

is approximately normally distributed with mean and variance

$$\mu_Z = \operatorname{arctanh} \rho = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} \qquad \text{and} \qquad \sigma_Z^2 = \frac{1}{n-3}$$

respectively. Therefore, to test the hypothesis H_0: $\rho = \rho_0$, we may use the test statistic

$$Z_0 = (\operatorname{arctanh} \hat{\rho} - \operatorname{arctanh} \rho_0)(n-3)^{1/2} \tag{11.49}$$

and reject H_0: $\rho = \rho_0$ if the value of the test statistic in Equation 11.49 is such that $|z_0| > z_{\alpha/2}$.

It is also possible to construct an approximate $100(1-\alpha)\%$ confidence interval for ρ using the transformation in Equation 11.48. The approximate $100(1-\alpha)\%$ confidence interval is

Confidence Interval for a Correlation Coefficient

$$\tanh\left(\operatorname{arctanh} \hat{\rho} - \frac{z_{\alpha/2}}{\sqrt{n-3}}\right) \leq \rho \leq \tanh\left(\operatorname{arctanh} \hat{\rho} + \frac{z_{\alpha/2}}{\sqrt{n-3}}\right) \tag{11.50}$$

where $\tanh u = (e^u - e^{-u})/(e^u + e^{-u})$.

EXAMPLE 11.8 | Wire Bond Pull Strength

Chapter 1 (Section 1.3) describes an application of regression analysis in which an engineer at a semiconductor assembly plant is investigating the relationship between pull strength of a wire bond and two factors: wire length and die height. In this example, we consider only one of the factors, the wire length. A random sample of 25 units is selected and tested, and the wire bond pull strength and wire length are observed for each unit. The data are shown in Table 1.2. We assume that pull strength and wire length are jointly normally distributed.

Figure 11.13 shows a scatter diagram of wire bond strength versus wire length. We have displayed box plots of each individual variable on the scatter diagram. There is

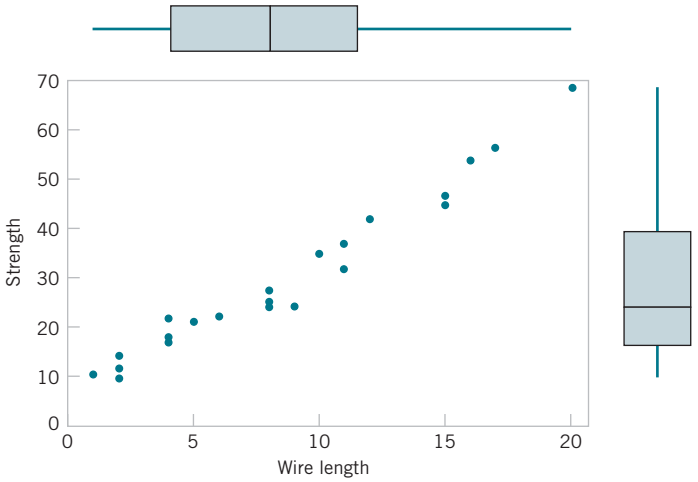


FIGURE 11.13

Scatter plot of wire bond strength versus wire length.

evidence of a linear relationship between the two variables. Typical computer output for fitting a simple linear regression model to the data is on the next page.

Now $S_{xx} = 698.56$ and $S_{xy} = 2027.7132$, and the sample correlation coefficient is

$$\hat{\rho} = \frac{S_{xy}}{[S_{xx}SS_T]^{1/2}} = \frac{2027.7132}{[(698.560)(6105.9)]^{1/2}} = 0.9818$$

Note that $\hat{\rho}_2 = (0.9818)^2 = 0.9640$ (which is reported in the computer output), or that approximately 96.40% of the variability in pull strength is explained by the linear relationship to wire length.

Now suppose that we wish to test the hypotheses

$$H_0: \rho = 0 \qquad H_1: \rho \neq 0$$

```
Strength = 5.11 + 2.90 Length
Predictor   Coef   SE Coef    T      P
Constant   5.115   1.146     4.46  0.000
Length     2.9027  0.1170   24.80  0.000
S = 3.093   R-sq = 96.4%  R-sq(adj) = 96.2%

PRESS = 272.144   R-sq(pred) = 95.54%

Analysis of Variance
Source       DF      SS      MS      F       P
Regression    1    5885.9  5885.9  615.08  0.000
Residual     23     220.1     9.6
  Error
Total        24    6105.9
```

with $\alpha = 0.05$. We can compute the t-statistic of Equation 11.46 as

$$t_0 = \frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}} = \frac{0.9818\sqrt{23}}{\sqrt{1-0.9640}} = 24.8$$

This statistic is also reported in the computer output as a test of $H_0: \beta_1 = 0$. Because $t_{0.025,23} = 2.069$, we reject H_0 and conclude that the correlation coefficient $\rho \neq 0$.

Finally, we may construct an approximate 95% confidence interval on ρ from Equation 11.50. Because arctanh $\hat{\rho}$ = arctanh 0.9818 = 2.3452, Equation 11.50 becomes

$$\tanh\left(2.3452 - \frac{1.96}{\sqrt{22}}\right) \leq \rho \leq \tanh\left(2.3452 + \frac{1.96}{\sqrt{22}}\right)$$

which reduces to

$$0.9585 \leq \rho \leq 0.9921$$

11.9 Regression on Transformed Variables

We occasionally find that the straight-line regression model $Y = \beta_0 + \beta_1 x + \varepsilon$ is inappropriate because the true regression function is nonlinear. Sometimes nonlinearity is visually determined from the scatter diagram, and sometimes, because of prior experience or underlying theory, we know in advance that the model is nonlinear. Occasionally, a scatter diagram will exhibit an apparent nonlinear relationship between Y and x. In some of these situations, a nonlinear function can be expressed as a straight line by using a suitable transformation. Such nonlinear models are called **intrinsically linear**.

As an example of a nonlinear model that is intrinsically linear, consider the exponential function

$$Y = \beta_0 e^{\beta_1 x} \varepsilon$$

This function is intrinsically linear because it can be transformed to a straight line by a logarithmic transformation

$$\ln Y = \ln \beta_0 + \beta_1 x + \ln \varepsilon$$

This transformation requires that the transformed error terms $\ln \varepsilon$ are normally and independently distributed with mean 0 and variance σ^2.

Another intrinsically linear function is

$$Y = \beta_0 + \beta_1 \left(\frac{1}{x}\right) + \varepsilon$$

By using the reciprocal transformation $z = 1/x$, the model is linearized to

$$Y = \beta_0 + \beta_1 z + \varepsilon$$

Sometimes several transformations can be employed jointly to linearize a function. For example, consider the function

$$Y = \frac{1}{\exp(\beta_0 + \beta_1 x + \varepsilon)}$$

letting $Y^* = 1/Y$, we have the linearized form

$$\ln Y^* = \beta_0 + \beta_1 x + \varepsilon$$

For examples of fitting these models, refer to Montgomery, Peck, and Vining (2012) or Myers (1990).

Transformations can be very useful in many situations in which the true relationship between the response Y and the regressor x is not well approximated by a straight line. The utility of a transformation is illustrated in the following example.

EXAMPLE 11.9 | Windmill Power

A research engineer is investigating the use of a windmill to generate electricity and has collected data on the DC output from this windmill and the corresponding wind velocity. The data are plotted in Figure 11.14 and listed in Table 11.5.

Inspection of the scatter diagram indicates that the relationship between DC output Y and wind velocity (x) may be nonlinear. However, we initially fit a straight-line model to the data. The regression model is

$$\hat{y} = 0.1309 + 0.2411\, x$$

The summary statistics for this model are $R^2 = 0.8745$, $MS_E = \hat{\sigma}^2 = 0.0557$, and $F_0 = 160.26$ (the P-value is <0.0001).

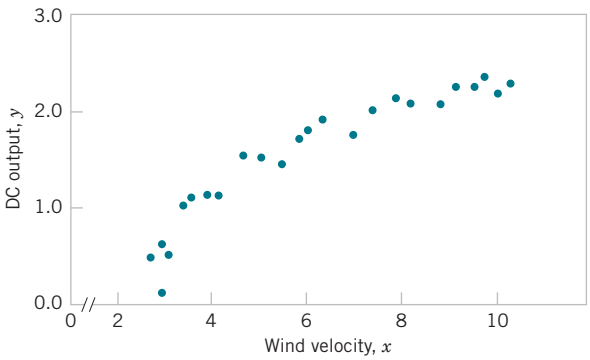


FIGURE 11.14
Plot of DC output y versus wind velocity x for the windmill data.

TABLE 11.5 Observed Values and Regressor Variable

Observation Number, i	Wind Velocity (mph), xi	DC Output, yi	Observation Number, i	Wind Velocity (mph), xi	DC Output, yi
1	5.00	1.582	14	5.80	1.737
2	6.00	1.822	15	7.40	2.088
3	3.40	1.057	16	3.60	1.137
4	2.70	0.500	17	7.85	2.179
5	10.00	2.236	18	8.80	2.112
6	9.70	2.386	19	7.00	1.800
7	9.55	2.294	20	5.45	1.501
8	3.05	0.558	21	9.10	2.303
9	8.15	2.166	22	10.20	2.310
10	6.20	1.866	23	4.10	1.194
11	2.90	0.653	24	3.95	1.144
12	6.35	1.930	25	2.45	0.123
13	4.60	1.562			

A plot of the residuals versus $\hat{y}_i$ is shown in Figure 11.15. This residual plot indicates model inadequacy and implies that the linear relationship has not captured all of the information in the wind speed variable. Note that the curvature that was apparent in the scatter diagram of Figure 11.14 is greatly amplified in the residual plots. Clearly, some other model form must be considered.

We might initially consider using a quadratic model such as

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

to account for the apparent curvature. However, the scatter diagram in Figure 11.14 suggests that as wind speed increases, DC output approaches an upper limit of approximately 2.5. This is also consistent with the theory of windmill operation. Because the quadratic model will eventually bend downward as wind speed increases, it would not be appropriate for these data. A more reasonable model for the windmill data that incorporates an upper asymptote would be

$$y = \beta_0 + \beta_1 \left(\frac{1}{x}\right) + \varepsilon$$

Figure 11.16 is a scatter diagram with the transformed variable $x' = 1/x$. This plot appears linear, indicating that the reciprocal transformation is appropriate. The fitted regression model is

$$\hat{y} = 2.9789 - 6.9345\, x'$$

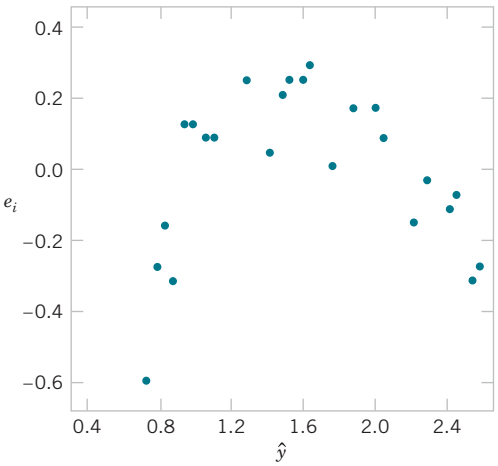


FIGURE 11.15
Plot of residuals e_i versus fitted values $\hat{y}_i$ for the windmill data.

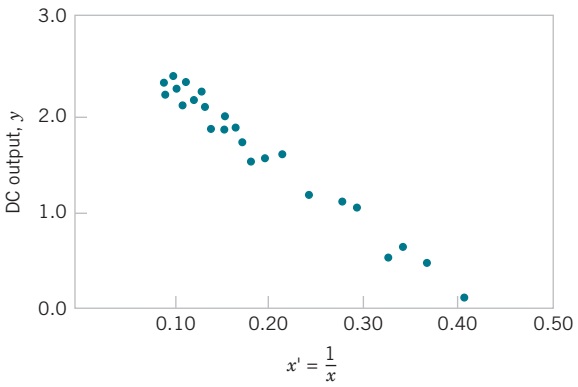


FIGURE 11.16
Plot of DC output versus $x' = 1/x$ for the windmill data.

The summary statistics for this model are $R^2 = 0.9800, MS_E = \hat{\sigma}^2 = 0.0089$, and $F_0 = 1128.43$ (the P-value is <0.0001).

A plot of the residuals from the transformed model versus $\hat{y}$ is shown in Figure 11.17. This plot does not reveal any serious problem with inequality of variance. The normal probability plot, shown in Figure 11.18, gives a mild indication that the errors come from a distribution with heavier tails than the normal (notice the slight upward and downward curve at the extremes). This normal probability plot has the z-score value plotted on the horizontal axis. Because there is no strong signal of model inadequacy, we conclude that the transformed model is satisfactory.

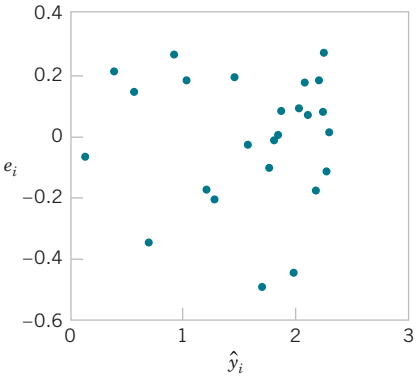


FIGURE 11.17

Plot of residuals versus fitted values $\hat{y}_i$ for the transformed model for the windmill data.

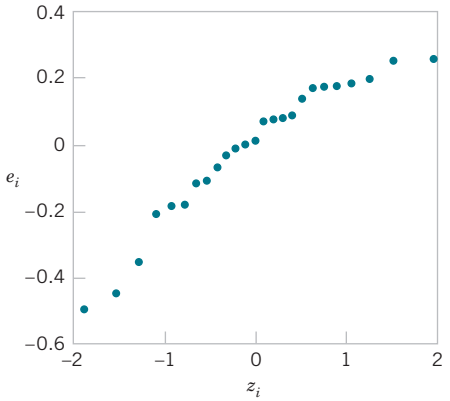


FIGURE 11.18

Normal probability plot of the residuals for the transformed model for the windmill data.

11.10 Logistic Regression

Linear regression often works very well when the response variable is **quantitative**. We now consider the situation in which the response variable takes on only two possible values, 0 and 1. These could be arbitrary assignments resulting from observing a **qualitative** response. For example, the response could be the outcome of a functional electrical test on a semiconductor device for which the results are either a "success," which means that the device works properly, or a "failure," which could be due to a short, an open, or some other functional problem.

Suppose that the model has the form

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{11.51}$$

and the response variable Y_i takes on the values either 0 or 1. We assume that the response variable Y_i is a **Bernoulli random variable** with probability distribution as follows:

Y_i	Probability
1	$P(Y_i = 1) = \pi_i$
0	$P(Y_i = 0) = 1 - \pi_i$

Now because $E(\varepsilon_i) = 0$, the expected value of the response variable is

$$E(Y_i) = 1(\pi_i) + 0(1 - \pi_i) = \pi_i$$

This implies that

$$E(Y_i) = \beta_0 + \beta_1 x_i = \pi_i$$

This means that the expected response given by the response function $E(Y_i) = \beta_0 + \beta_1 x_i$ is just the probability that the response variable takes on the value 1.

There are some substantive problems with the regression model in Equation 11.51. First, note that if the response is binary, the error terms ε_i can only take on two values, namely,

$$\varepsilon_i = 1 - (\beta_0 + \beta_1 x_i) \text{ when } Y_i = 1$$
$$\varepsilon_i = -(\beta_0 + \beta_1 x_i) \text{ when } Y_i = 0$$

Consequently, the errors in this model cannot possibly be normal. Second, the error variance is not constant, because

$$\begin{aligned}\sigma^2_{Y_i} &= E[Y_i - (Y_i)]^2 \\ &= (1 - \pi_i)^2\pi_i + (0 - \pi_i)^2(1 - \pi_i) \\ &= \pi_i(1 - \pi_i)\end{aligned}$$

Notice that this last expression is just

$$\sigma^2_{y_i} = E(Y_i)[1 - E(Y_i)]$$

because $E(Y_i) = \beta_0 + \beta_1 x_i = \pi_i$. This indicates that the variance of the observations (which is the same as the variance of the errors because $\varepsilon_i = Y_i - \pi_i$, and π_i is a constant) is a function of the mean. Finally, there is a constraint on the response function because

$$0 \le E(Y_i) = \pi_i \le 1$$

This restriction can cause serious problems with the choice of a **linear response function** as we have initially assumed in Equation 11.51. It would be possible to fit a model to the data for which the predicted values of the response lie outside the 0, 1 interval.

Generally, when the response variable is binary, there is considerable empirical evidence indicating that the shape of the response function should be nonlinear. A monotonically increasing (or decreasing) *S*-shaped (or reverse *S*-shaped) function, such as that shown in Figure 11.19, is usually employed. This function is called the **logit response function**, and has the form

$$E(Y) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \tag{11.52}$$

or equivalently,

$$E(Y) = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 x)]} \tag{11.53}$$

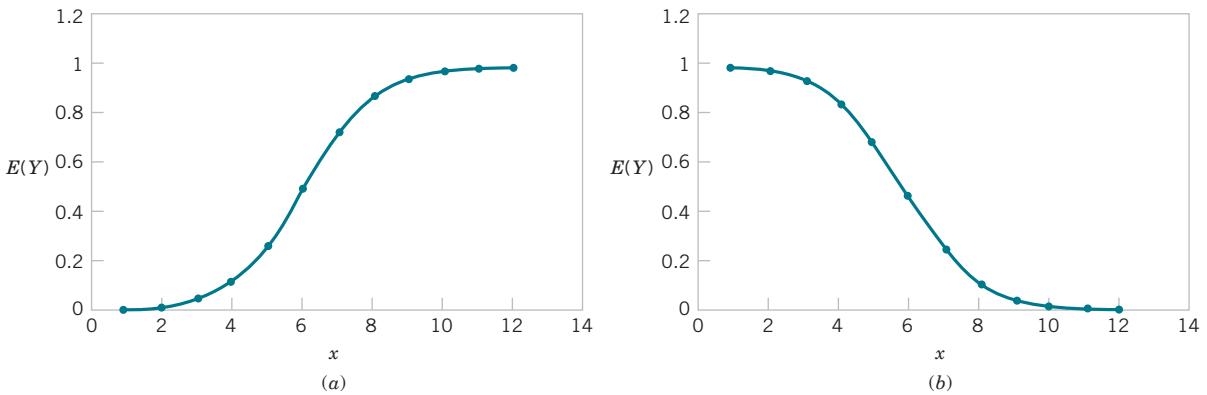


FIGURE 11.19 **Examples of the logistic response function. (a) $E(Y) = 1/(1 + e^{-6.0 - 1.0x})$. (b) $E(Y) = 1/(1 + e^{-6.0 + 1.0x})$.**

In **logistic regression**, we assume that $E(Y)$ is related to x by the logit function. It is easy to show that

$$\frac{E(Y)}{1 - E(Y)} = \exp(\beta_0 + \beta_1 x) \tag{11.54}$$

The quantity in Equation 11.54 is called the **odds**. It has a straightforward interpretation: If the odds is 2 for a particular value of x, it means that a success is twice as likely as a failure at that value of the regressor x. Notice that the natural logarithm of the odds is a linear function of the regressor variable. Therefore, the slope β_1 is the difference in the log odds that results from a one-unit increase in x. This means that the odds ratio equals e^{β_1} when x increases by one unit.

The parameters in this logistic regression model are usually estimated by the method of maximum likelihood. For details of the procedure, see Montgomery, Peck, and Vining (2012). Computer software will fit logistic regression models and provide useful information on the quality of the fit.

We illustrate logistic regression using the data on launch temperature and O-ring failure for the 24 Space Shuttle launches prior to the *Challenger* disaster of January 1986. Six O-rings were used to seal field joints on the rocket motor assembly. The following table presents the launch temperatures. A "1" in the "O-Ring Failure" column indicates that at least one O-ring failure had occurred on that launch.

Temperature	O-Ring Failure	Temperature	O-Ring Failure	Temperature	O-Ring Failure
53	1	68	0	75	0
56	1	69	0	75	1
57	1	70	0	76	0
63	0	70	1	76	0
66	0	70	1	78	0
67	0	70	1	79	0
67	0	72	0	80	0
67	0	73	0	81	0

Figure 11.20 is a scatter plot of the data. Note that failures tend to occur at lower temperatures. The logistic regression model fit to these data from a computer software package is shown in the following boxed display. (Both Minitab and JMP have excellent capability to fit logistic regression models.)

The fitted logistic regression model is

$$\hat{y} = \frac{1}{1 + \exp[-(10.875 - 0.17132x)]}$$

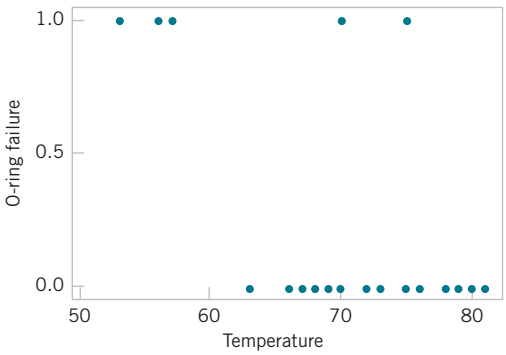


FIGURE 11.20
Scatter plot of O-ring failures versus launch temperature for 24 Space Shuttle flights.

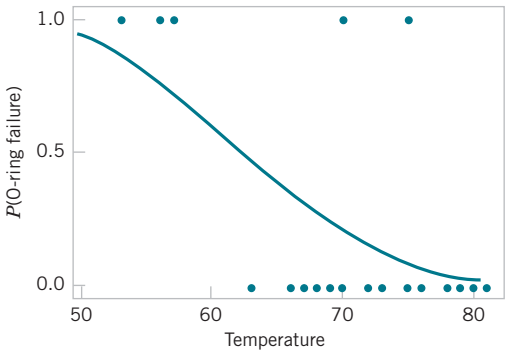


FIGURE 11.21

Probability of O-ring failure versus launch temperature (based on a logistic regression model).

The standard error of the slope $\hat{\beta}_1$ is $se(\hat{\beta}_1) = 0.08344$. For large samples, $\hat{\beta}_1$ has an approximate normal distribution, and so $\hat{\beta}_1/se(\hat{\beta}_1)$ can be compared to the standard normal distribution to test H_0: $\beta_1 = 0$. Software performs this test. The P-value is 0.04, indicating that temperature has a significant effect on the probability of O-ring failure. The odds ratio is 0.84, so every 1-degree increase in temperature reduces the odds of failure by 0.84. Figure 11.21 shows the fitted logistic regression model. The sharp increase in the probability of O-ring failure is very evident in this graph. The actual temperature at the *Challenger* launch was 31°F. This is well outside the range of other launch temperatures, so our logistic regression model is not likely to provide highly accurate predictions at that temperature, but it is clear that a launch at 31°F is almost certainly going to result in O-ring failure.

Binary Logistic Regression: O-Ring Failure versus Temperature

```
Link Function: Logit

Response Information
Variable    Value    Count
O-Ring F      1        7    (Event)
              0       17
            Total     24

Logistic Regression Table
                                                 Odds   95%           CI
Predictor    Coef    SE Coef     Z       P     Ratio  Lower         Upper
Constant   10.875    5.703      1.91   0.057
Temperat  -0.17132  0.08344   -2.05   0.040   0.84   0.72          0.99

Log-Likelihood = −11.515

Test that all slopes are zero: G = 5.944, DF = 1, P-Value = 0.015
```

It is interesting to note that all of these data were available **prior** to launch. However, engineers were unable to effectively analyze the data and use them to provide a convincing argument against launching *Challenger* to NASA managers. Yet a simple regression analysis of the data would have provided a strong quantitative basis for this argument. This is one of the more dramatic instances that points out *why engineers and scientists need a strong background in basic statistical techniques*.

Important Terms and Concepts

Analysis of variance table
Coefficient of determination
Confidence interval on the intercept
Confidence interval on the mean response
Confidence interval on the slope
Correlation coefficient
Empirical model
Error sum of squares
Intrinsically linear model
Least squares
Logistic regression
Logit response function
Mean squares
Normal probability plot of residuals
Odds ratio
Odds
Outlier
Prediction interval on a future observation
Regression analysis
Regression coefficients
Regression line
Regression sum of squares
Regressor
Residual plots
Residuals
Response variable
Significance of regression
Simple linear regression model standard errors
Statistical tests on model parameters
Total corrected sum of squares
Transformations

CHAPTER **12**

Multiple Linear Regression

© skynesher / iStockphoto

LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Use multiple regression techniques to build empirical models to engineering and scientific data
2. Understand how the method of least squares extends to fitting multiple regression models
3. Assess regression model adequacy
4. Test hypotheses and construct confidence intervals on the regression coefficients
5. Use the regression model to estimate the mean response and to make predictions and to construct confidence intervals and prediction intervals
6. Build regression models with polynomial terms
7. Use indicator variables to model categorical regressors
8. Use stepwise regression and other model-building techniques to select the appropriate set of variables for a regression model

CHAPTER OUTLINE

12.1 Multiple Linear Regression Model
- 12.1.1 Introduction
- 12.1.2 Least Squares Estimation of the Parameters
- 12.1.3 Matrix Approach to Multiple Linear Regression
- 12.1.4 Properties of the Least Squares Estimators

12.2 Hypothesis Tests in Multiple Linear Regression
- 12.2.1 Test for Significance of Regression
- 12.2.2 Tests on Individual Regression Coefficients and Subsets of Coefficients

12.3 Confidence Intervals in Multiple Linear Regression
- 12.3.1 Confidence Intervals on Individual Regression Coefficients
- 12.3.2 Confidence Interval on the Mean Response

12.4 Prediction of New Observations

12.5 Model Adequacy Checking
- 12.5.1 Residual Analysis
- 12.5.2 Influential Observations

12.6 Aspects of Multiple Regression Modeling
- 12.6.1 Polynomial Regression Models
- 12.6.2 Categorical Regressors and Indicator Variables
- 12.6.3 Selection of Variables and Model Building
- 12.6.4 Multicollinearity

This chapter generalizes the simple linear regression to a situation that has more than one predictor or regressor variable. This situation occurs frequently in science and engineering; for example, in Chapter 1, we provided data on the pull strength of a wire bond on a semiconductor package and illustrated its relationship to the wire length and the die height. Understanding the relationship between strength and the other two variables may provide important insight to the engineer when the package is designed, or to the manufacturing personnel who assemble the die into the package. We used a multiple **linear regression** model to relate strength to wire length and die height. There are many examples of such relationships: The life of a cutting tool is related to the cutting speed and the tool angle; patient satisfaction in a hospital is related to patient age, type of procedure performed, and length of stay; and the fuel economy of a vehicle is related to the type of vehicle (car versus truck), engine displacement, horsepower, type of transmission, and vehicle weight. Multiple regression models give insight into the relationships between these variables that can have important practical implications.

In this chapter, we show how to fit multiple linear regression models, perform the statistical tests and confidence procedures that are analogous to those for simple linear regression, and check for model adequacy. We also show how models that have polynomial terms in the regressor variables are just multiple linear regression models. We also discuss some aspects of building a good regression model from a collection of candidate regressors.

12.1 Multiple Linear Regression Model

12.1.1 Introduction

Many applications of regression analysis involve situations that have more than one regressor or predictor variable. A regression model that contains more than one regressor variable is called a **multiple regression model**.

As an example, suppose that the gasoline mileage performance of a vehicle depends on the vehicle weight and the engine displacement. A multiple regression model that might describe this relationship is

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \tag{12.1}$$

where Y represents the mileage, x_1 represents the weight, x_2 represents the engine displacement, and ϵ is a random error term. This is a multiple linear regression model with two regressors.

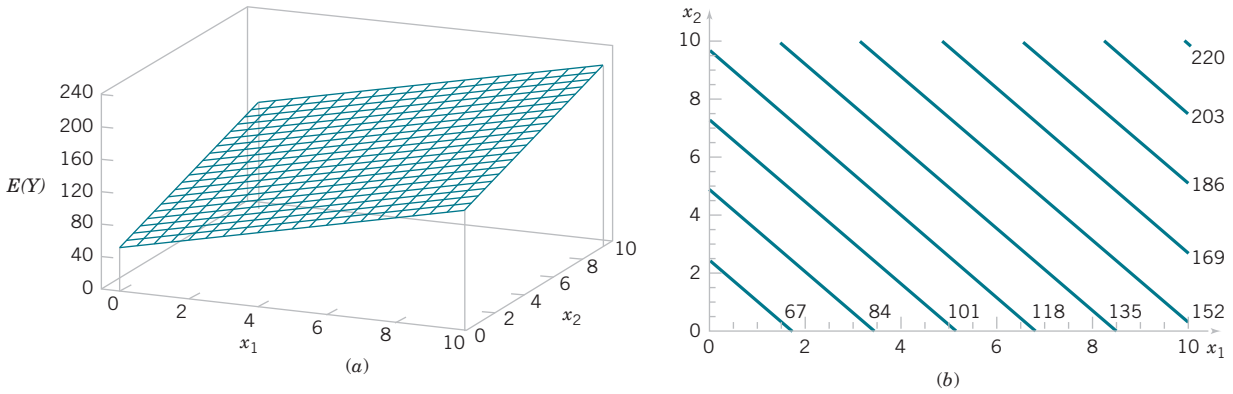


FIGURE 12.1

(a) The regression plane for the model $E(Y) = 50 + 10x_1 + 7x_2$. (b) The contour plot.

The term *linear* is used because Equation 12.1 is a linear function of the unknown parameters β_0, β_1, and β_2.

The regression model in Equation 12.1 describes a plane in the three-dimensional space of Y, x_1, and x_2. Figure 12.1(a) shows this plane for the regression model

$$E(Y) = 50 + 10x_1 + 7x_2$$

where we have assumed that the expected value of the error term is zero; that is, $E(\epsilon) = 0$. The parameter β_0 is the **intercept** of the plane. We sometimes call β_1 and β_2 **partial regression coefficients** because β_1 measures the expected change in Y per unit change in x_1 when x_2 is held constant, and β_2 measures the expected change in Y per unit change in x_2 when x_1 is held constant. Figure 12.1(b) shows a contour plot of the regression model—that is, lines of constant $E(Y)$ as a function of x_1 and x_2. Notice that the contour lines in this plot are straight lines.

In general, the dependent variable or response Y may be related to k independent or regressor variables. The model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \tag{12.2}$$

is called a *multiple linear regression model with k regressor variables*. The parameters β_j, $j = 0, 1, \ldots, k$, are called the *regression coefficients*. This model describes a hyperplane in the k+1-dimensional space of Y and the regressor variables $\{x_j\}$. The parameter β_j represents the expected change in response Y per unit change in x_j when all the remaining regressors $x_i (i \neq j)$ are held constant.

Multiple linear regression models are often used as approximating functions. That is, the true functional relationship between Y and $x_1, x_2, \ldots, x_k$ is unknown, but over certain ranges of the independent variables, the linear regression model is an adequate approximation.

Models that are more complex in structure than Equation 12.2 may often still be analyzed by multiple linear regression techniques. For example, consider the cubic polynomial model in one regressor variable.

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon \tag{12.3}$$

If we let $x_1 = x$, $x_2 = x^2$, $x_3 = x^3$, Equation 12.3 can be written as

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon \tag{12.4}$$

which is a multiple linear regression model with three regressor variables.

Models that include **interaction** effects may also be analyzed by multiple linear regression methods. Interaction effects are very common. For example, a vehicle's mileage may be impacted by an interaction between vehicle weight and engine displacement. An interaction between two variables can be represented by a cross-product term in the model, such as

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon \tag{12.5}$$

If we let $x_3 = x_1x_2$ and $\beta_3 = \beta_{12}$, Equation 12.5 can be written as

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \epsilon$$

which is a linear regression model.

Figure 12.2(a) and (b) shows the three-dimensional plot of the regression model

$$Y = 50 + 10x_1 + 7x_2 + 5x_1x_2$$

and the corresponding two-dimensional contour plot. Notice that, although this model is a linear regression model, the shape of the surface that is generated by the model is not linear. In general, **any regression model that is linear in parameters** (the β's) **is a linear regression model, regardless of the shape of the surface that it generates**.

Figure 12.2 provides a nice graphical interpretation of an interaction. Generally, interaction implies that the effect produced by changing one variable (x_1, say) depends on the level of the other variable (x_2). For example, Figure 12.2 shows that changing x_1 from 2 to 8 produces a much smaller change in $E(Y)$ when $x_2 = 2$ than when $x_2 = 10$. Interaction effects occur frequently in the study and analysis of real-world systems, and regression methods are one of the techniques that we can use to describe them.

As a final example, consider the second-order model with interaction

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \beta_{12}x_1x_2 + \epsilon \tag{12.6}$$

If we let $x_3 = x_1^2, x_4 = x_2^2, x_5 = x_1x_2, \beta_3 = \beta_{11}, \beta_4 = \beta_{22}$, and $\beta_5 = \beta_{12}$, Equation 12.6 can be written as a multiple linear regression model as follows:

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 + \epsilon$$

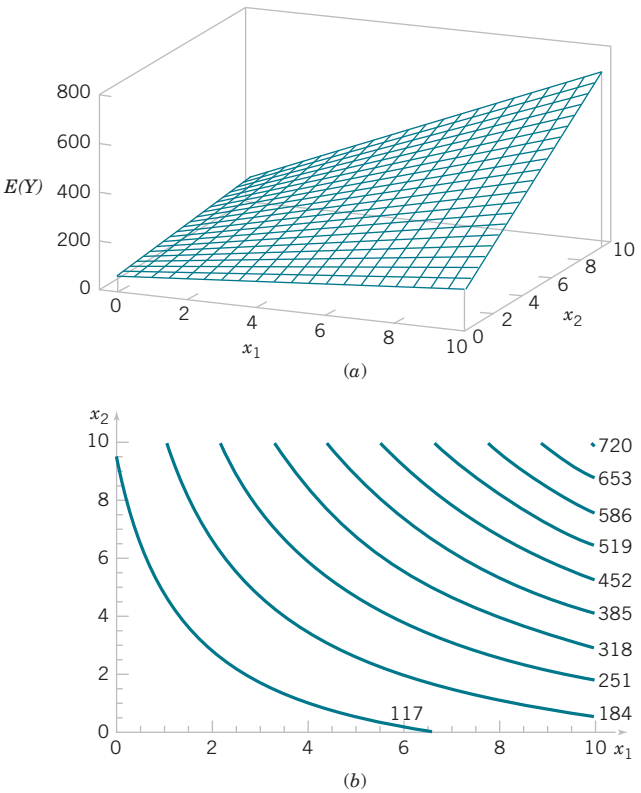


FIGURE 12.2

(a) Three-dimensional plot of the regression model. (b) The contour plot.

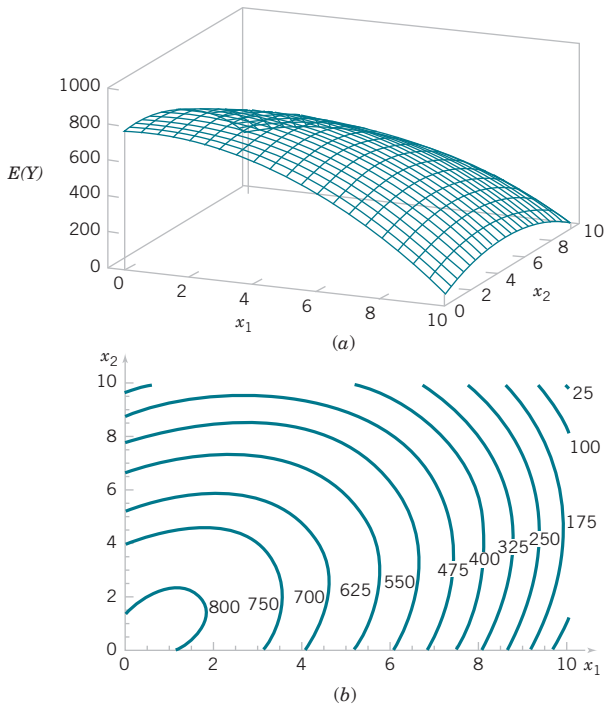


FIGURE 12.3

(a) Three-dimensional plot of the regression model $E(Y) = 800 + 10x_1 + 7x_2 - 8.5x_1^2 - 5x_2^2 + 4x_1x_2$. (b) The contour plot.

Figure 12.3 parts (a) and (b) show the three-dimensional plot and the corresponding contour plot for

$$E(Y) = 800 + 10x_1 + 7x_2 - 8.5x_1^2 - 5x_2^2 + 4x_1x_2$$

These plots indicate that the expected change in Y when x_i is changed by one unit (say) is a function of and x_2. The quadratic and interaction terms in this model produce a mound-shaped function. Depending on the values of the regression coefficients, the second-order model with interaction is capable of assuming a wide variety of shapes; thus, it is a very flexible regression model.

12.1.2 Least Squares Estimation of the Parameters

The method of least squares may be used to estimate the regression coefficients in the multiple regression model, Equation 12.2. Suppose that $n > k$ observations are available, and let x_{ij} denote the ith observation or level of variable x_j. The observations are

$$(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i), \quad i = 1, 2, \ldots, n \quad \text{and} \quad n > k$$

It is customary to present the data for multiple regression in a table such as Table 12.1.

Each observation $(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i)$, satisfies the model in Equation 12.2, or

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \epsilon_i \\ &= \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \epsilon_i \qquad i = 1, 2, \ldots, n \end{aligned} \tag{12.7}$$

The least squares function is

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2 \tag{12.8}$$

We want to minimize L with respect to $\beta_0, \beta_1, \ldots, \beta_k$. The **least squares estimates** of $\beta_0, \beta_1, \ldots, \beta_k$ must satisfy

$$\left. \frac{\partial L}{\partial \beta_j} \right|_{\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k} = -2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{k} \hat{\beta}_j x_{ij} \right) = 0 \tag{12.9a}$$

and

$$\left. \frac{\partial L}{\partial \beta_j} \right|_{\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k} = -2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{k} \hat{\beta}_j x_{ij} \right) x_{ij} = 0 \quad j = 1, 2, \ldots, k \tag{12.9b}$$

TABLE 12.1 Data for Multiple Linear Regression

y	x_1	x_2	$\cdots$	x_k
y_1	x_{11}	x_{12}	$\cdots$	x_{1k}
y_2	x_{21}	x_{22}	$\cdots$	x_{2k}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
y_n	x_{n1}	x_{n2}	$\cdots$	x_{nk}

Simplifying Equation 12-9, we obtain the **least squares normal equations**

$$\begin{aligned}
n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik} &= \sum_{i=1}^{n} y_i \\
\hat{\beta}_0 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^{n} x_{i1}x_{i2} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} x_{i1}x_{ik} &= \sum_{i=1}^{n} x_{i1}y_i \\
&\vdots \\
\hat{\beta}_0 \sum_{i=1}^{n} x_{ik} + \hat{\beta}_1 \sum_{i=1}^{n} x_{ik}x_{i1} + \hat{\beta}_2 \sum_{i=1}^{n} x_{ik}x_{i2} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik}^2 &= \sum_{i=1}^{n} x_{ik}y_i
\end{aligned} \tag{12.10}$$

Note that there are $p = k + 1$ normal equations, one for each of the unknown regression coefficients. The solution to the normal equations will be the least squares estimators of the regression coefficients, $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$. The normal equations can be solved by any method appropriate for solving a system of linear equations.

EXAMPLE 12.1 | Wire Bond Strength

In Chapter 1, we used data on pull strength of a wire bond in a semiconductor manufacturing process, wire length, and die height to illustrate building an empirical model. We use the same data, repeated for convenience in Table 12.2, and show the details of estimating the model parameters. A three-dimensional scatter plot of the data is presented in Figure 1.15. Figure 12.4 is a matrix of two-dimensional scatter plots of the data. These displays can be helpful in visualizing the relationships among variables in a multivariable data set. For example, the plot indicates that there is a strong linear relationship between strength and wire length.

Specifically, we fit the multiple linear regression model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

where Y= pull strength, x_1 = wire length, and x_2 = die height. From the data in Table 12.2, we calculate

$$n = 25, \sum_{i=1}^{25} y_i = 725.82, \sum_{i=1}^{25} x_{i1} = 206, \sum_{i=1}^{25} x_{i2} = 8{,}294$$

$$\sum_{i=1}^{25} x_{i1}^2 = 2{,}396, \sum_{i=1}^{25} x_{i2}^2 = 3{,}531{,}848, \sum_{i=1}^{25} x_{i1}x_{i2} = 77{,}177$$

$$\sum_{i=1}^{25} x_{i1}y_i = 8{,}008.47, \sum_{i=1}^{25} x_{i2}y_i = 274{,}816.71$$

TABLE 12.2 Wire Bond Data

Observation Number	Pull Strength y	Wire Length x_1	Die Height x_2	Observation Number	Pull Strength y	Wire Length x_1	Die Height x_2
1	9.95	2	50	14	11.66	2	360
2	24.45	8	110	15	21.65	4	205
3	31.75	11	120	16	17.89	4	400
4	35.00	10	550	17	69.00	20	600
5	25.02	8	295	18	10.30	1	585
6	16.86	4	200	19	34.93	10	540
7	14.38	2	375	20	46.59	15	250
8	9.60	2	52	21	44.88	15	290
9	24.35	9	100	22	54.12	16	510
10	27.50	8	300	23	56.63	17	590
11	17.08	4	412	24	22.13	6	100
12	37.00	11	400	25	21.15	5	400
13	41.95	12	500				

For the model $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$, the normal Equations 12.10 are

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} y_{i1}$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^{n} x_{i1}x_{i2} = \sum_{i=1}^{n} x_{i1}y_{i1}$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{i2} + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1}x_{i2} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2}^2 = \sum_{i=1}^{n} x_{i2}y_i$$

Inserting the computed summations into the normal equations, we obtain

$$\begin{aligned} 25\hat{\beta}_0 + \quad 206\hat{\beta}_1 + \quad 8294\hat{\beta}_2 &= 725.82 \\ 206\hat{\beta}_0 + \quad 2396\hat{\beta}_1 + \quad 77{,}177\hat{\beta}_2 &= 8{,}008.47 \\ 8294\hat{\beta}_0 + 77{,}177\hat{\beta}_1 + 3{,}531{,}848\hat{\beta}_2 &= 274{,}816.71 \end{aligned}$$

The solution to this set of equations is

$$\hat{\beta}_0 = 2.26379, \quad \hat{\beta}_1 = 2.74427, \quad \hat{\beta}_2 = 0.01253$$

Therefore, the fitted regression equation is

$$\hat{y} = 2.26379 + 2.74427x_1 + 0.01253x_2$$

Practical Interpretation: This equation can be used to predict pull strength for pairs of values of the regressor variables wire length (x_1) and die height (x_2). This is essentially the same regression model given in Section 1.3. Figure 1.16 shows a three-dimensional plot of the plane of predicted values $\hat{y}$ generated from this equation.

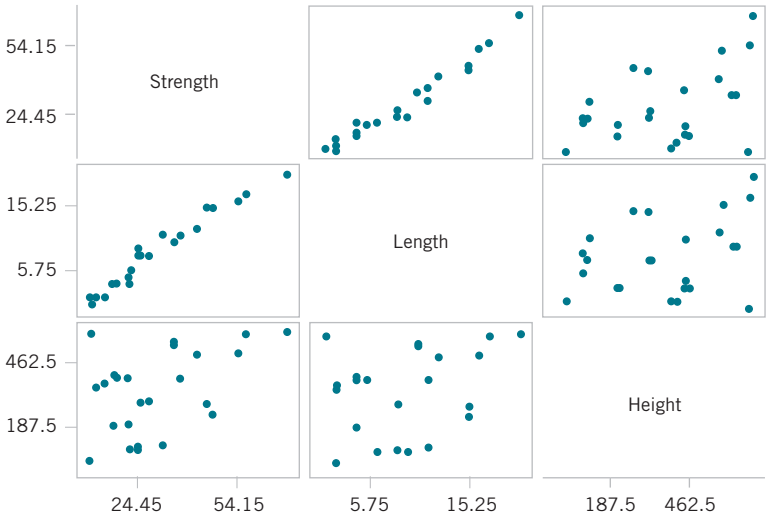


FIGURE 12.4

Matrix of computer-generated scatter plots for the wire bond pull strength data in Table 12.2.

12.1.3 Matrix Approach to Multiple Linear Regression

In fitting a multiple regression model, it is much more convenient to express the mathematical operations using **matrix notation**. Suppose that there are k regressor varibles and n observations, $(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i)$, $i = 1, 2, \ldots, n$ and that the model relating the regressors to the response is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \epsilon_i \quad i = 1, 2, \ldots, n$$

This model is a system of n equations that can be expressed in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{12.11}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad \text{and} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

In general, $\mathbf{y}$ is an $(n \times 1)$ vector of the observations, $\mathbf{X}$ is an $(n \times p)$ matrix of the levels of the independent variables (assuming that the intercept is always multiplied by a constant value—unity), $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of the regression coefficients, and $\boldsymbol{\epsilon}$ is a $(n \times 1)$ vector of random errors. The $\mathbf{X}$ matrix is often called the **model matrix**.

We wish to find the vector of least squares estimators, $\hat{\boldsymbol{\beta}}$, that minimizes

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

The least squares estimator $\hat{\boldsymbol{\beta}}$ is the solution for $\boldsymbol{\beta}$ in the equations

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = 0$$

We will not give the details of taking the preceding derivatives; however, the resulting equations that must be solved are

Normal Equations

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \tag{12.12}$$

Equations 12.12 are the least squares normal equations in matrix form. They are identical to the scalar form of the normal equations given earlier in Equations 12.10. To solve the normal equations, multiply both sides of Equations 12.12 by the inverse of $\mathbf{X}'\mathbf{X}$. Therefore, the least squares estimate of $\boldsymbol{\beta}$ is

Least Squares Estimate of β

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{12.13}$$

Note that there are $p = k + 1$ normal equations in $p = k + 1$ unknowns (the values of $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$). Furthermore, the matrix $\mathbf{X}'\mathbf{X}$ is always nonsingular, as was assumed previously, so the methods described in textbooks on determinants and matrices for inverting these matrices can be used to find $(\mathbf{X}'\mathbf{X})^{-1}$. In practice, multiple regression calculations are almost always performed using a computer.

It is easy to see that the matrix form of the normal equations is identical to the scalar form. Writing out Equation 12.12 in detail, we obtain

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \cdots & \sum_{i=1}^{n} x_{ik} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1}x_{i2} & \cdots & \sum_{i=1}^{n} x_{i1}x_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} x_{ik} & \sum_{i=1}^{n} x_{ik}x_{i1} & \sum_{i=1}^{n} x_{ik}x_{i2} & \cdots & \sum_{i=1}^{n} x_{ik}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1}y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ik}y_i \end{bmatrix}$$

If the indicated matrix multiplication is performed, the scalar form of the normal equations (that is, Equation 12.10) will result. In this form, it is easy to see that $\mathbf{X}'\mathbf{X}$ is a $(p \times p)$ symmetric matrix and $\mathbf{X}'\mathbf{y}$ is a $(p \times 1)$ column vector. Note the special structure of the $\mathbf{X}'\mathbf{X}$ matrix.

The diagonal elements of $\mathbf{X'X}$ are the sums of squares of the elements in the columns of $\mathbf{X}$, and the off-diagonal elements are the sums of cross-products of the elements in the columns of X. Furthermore, note that the elements of $\mathbf{X'y}$ are the sums of cross-products of the columns of $\mathbf{X}$ and the observations $\{y_i\}$.

The fitted regression model is

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j x_{ij} \qquad i = 1, 2, \ldots, n \tag{12.14}$$

In matrix notation, the fitted model is

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

The difference between the observation y_i and the fitted value $\hat{y}_i$ is a **residual**, say, $e_i = y_i - \hat{y}_i$. The $(n \times 1)$ vector of residuals is denoted by

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} \tag{12.15}$$

EXAMPLE 12.2 | Wire Bond Strength with Matrix Notation

In Example 12.1, we illustrated fitting the multiple regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

where y is the observed pull strength for a wire bond, x_1 is the wire length, and x_2 is the die height. The 25 observations are in Table 12.2. We will now use the matrix approach to fit the previous regression model to these data. The model matrix $\mathbf{X}$ and $\mathbf{y}$ vector for this model are

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 50 \\ 1 & 8 & 110 \\ 1 & 11 & 120 \\ 1 & 10 & 550 \\ 1 & 8 & 295 \\ 1 & 4 & 200 \\ 1 & 2 & 375 \\ 1 & 2 & 52 \\ 1 & 9 & 100 \\ 1 & 8 & 300 \\ 1 & 4 & 412 \\ 1 & 11 & 400 \\ 1 & 12 & 500 \\ 1 & 2 & 360 \\ 1 & 4 & 205 \\ 1 & 4 & 400 \\ 1 & 20 & 600 \\ 1 & 1 & 585 \\ 1 & 10 & 540 \\ 1 & 15 & 250 \\ 1 & 15 & 290 \\ 1 & 16 & 510 \\ 1 & 17 & 590 \\ 1 & 6 & 100 \\ 1 & 5 & 400 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} 9.95 \\ 24.45 \\ 31.75 \\ 35.00 \\ 25.02 \\ 16.86 \\ 14.38 \\ 9.60 \\ 24.35 \\ 27.50 \\ 17.08 \\ 37.00 \\ 41.95 \\ 11.66 \\ 21.65 \\ 17.89 \\ 69.00 \\ 10.30 \\ 34.93 \\ 46.59 \\ 44.88 \\ 54.12 \\ 56.63 \\ 22.13 \\ 21.15 \end{bmatrix}$$

The $\mathbf{X'X}$ matrix is

$$\mathbf{X'X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 2 & 8 & \cdots & 5 \\ 50 & 110 & \cdots & 400 \end{bmatrix} \begin{bmatrix} 1 & 2 & 50 \\ 1 & 8 & 110 \\ \vdots & \vdots & \vdots \\ 1 & 5 & 400 \end{bmatrix} = \begin{bmatrix} 25 & 206 & 8{,}294 \\ 206 & 2{,}396 & 77{,}177 \\ 8{,}294 & 77{,}177 & 3{,}531{,}848 \end{bmatrix}$$

and the $\mathbf{X'y}$ vector is

$$\mathbf{X'y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 2 & 8 & \cdots & 5 \\ 50 & 110 & \cdots & 400 \end{bmatrix} \begin{bmatrix} 9.95 \\ 24.45 \\ \vdots \\ 21.15 \end{bmatrix} = \begin{bmatrix} 725.82 \\ 8{,}008.47 \\ 274{,}816.71 \end{bmatrix}$$

The least squares estimates are found from Equation 12.13 as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y}$$

or

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 25 & 206 & 8{,}294 \\ 206 & 2{,}396 & 77{,}177 \\ 8{,}294 & 77{,}177 & 3{,}531{,}848 \end{bmatrix}^{-1} \begin{bmatrix} 725.82 \\ 8{,}008.37 \\ 274{,}11.31 \end{bmatrix}$$

$$= \begin{bmatrix} 0.214653 & -0.007491 & -0.000340 \\ -0.007491 & 0.001671 & -0.000019 \\ -0.000340 & -0.000019 & +0.0000015 \end{bmatrix} \begin{bmatrix} 725.82 \\ 8{,}008.47 \\ 274{,}811.31 \end{bmatrix}$$

$$= \begin{bmatrix} 2.26379143 \\ 2.74426964 \\ 0.01252781 \end{bmatrix}$$

Therefore, the fitted regression model with the regression coefficients rounded to five decimal places is

$$\hat{y}_1 = 2.26379 + 2.74427x_1 + 0.01253x_2$$

This is identical to the results obtained in Example 12.1.

This regression model can be used to predict values of pull strength for various values of wire length (x_1) and die height (x_2). We can also obtain the *fitted values* $\hat{y}_i$ by substituting each observation (x_{i1}, x_{i2}), $i = 1, 2, \ldots, n$, into the equation. For example, the first observation has $x_{11} = 2$ and $x_{12} = 50$, and the fitted value is

$$\begin{aligned}\hat{y}_1 &= 2.26379 + 2.74427x_{11} + 0.01253x_{12}\\ &= 2.26379 + 2.74427(2) + 0.01253(50) = 8.38\end{aligned}$$

The corresponding observed value is $y_1 = 9.95$. The *residual* corresponding to the first observation is

$$e_1 = y_1 - \hat{y}_1 = 9.95 - 8.38 = 1.57$$

Table 12.3 displays all 25 fitted values $\hat{y}_i$ and the corresponding residuals. The fitted values and residuals are calculated to the same accuracy as the original data.

TABLE 12.3 Observations, Fitted Values, and Residuals

Observation Number	y_i	$\hat{y}_i$	$e_i = y_i - \hat{y}_i$	Observation Number	y_i	$\hat{y}_i$	$e_i = y_i - \hat{y}_i$
1	9.95	8.38	1.57	14	11.66	12.26	−0.60
2	24.45	25.60	−1.15	15	21.65	15.81	5.84
3	31.75	33.95	−2.20	16	17.89	18.25	−0.36
4	35.00	36.60	−1.60	17	69.00	64.67	4.33
5	25.02	27.91	−2.89	18	10.30	12.34	−2.04
6	16.86	15.75	1.11	19	34.93	36.47	−1.54
7	14.38	12.45	1.93	20	46.59	46.56	0.03
8	9.60	8.40	1.20	21	44.88	47.06	−2.18
9	24.35	28.21	−3.86	22	54.12	52.56	1.56
10	27.50	27.98	−0.48	23	56.63	56.31	0.32
11	17.08	18.40	−1.32	24	22.13	19.98	2.15
12	37.00	37.46	−0.46	25	21.15	21.00	0.15
13	41.95	41.46	0.49				

Computers are almost always used in fitting multiple regression models. See Table 12.4 for some annotated computer output for the least squares regression model for the wire bond pull strength data. The upper part of the table contains the numerical estimates of the regression coefficients. The computer also calculates several other quantities that reflect important information about the regression model. In subsequent sections, we define and explain the quantities in this output.

Estimating σ^2 Just as in simple linear regression, it is important to estimate σ^2, the variance of the error term ϵ, in a multiple regression model. Recall that in simple linear regression the estimate of σ^2 was obtained by dividing the sum of the squared residuals by $n - 2$. Now there

TABLE 12.4 **Multiple Regression Output from Software for the Wire Bond Pull Strength Data**

Regression Analysis: Strength versus Length, Height

Strength = 2.26 + 2.74 Length + 0.0125 Height

Predictor	Coef	SE Coef	T	P	VIF
Constant	$\hat{\beta}_0$ → 2.264	1.060	2.14	0.044	
Length	$\hat{\beta}_1$ → 2.74427	0.09352	29.34	0.000	1.2
Height	$\hat{\beta}_2$ → 0.012528	0.002798	4.48	0.000	1.2

S = 2.288	R-Sq = 98.1%	R-Sq (adj) = 97.9%
PRESS = 156.163	R-Sq (pred) = 97.44%	

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	5990.8	2995.4	572.17	0.000
Residual error	22	115.2	5.2 ← $\hat{\sigma}^2$		
Total	24	6105.9			

Source	DF	Seq SS
Length	1	5885.9
Height	1	104.9

Predicted Values for New Observations

New Obs	Fit	SE Fit	95.0% CI	95.0% PI
1	27.663	0.482	(26.663, 28.663)	(22.814, 32.512)

Values of Predictors for New Observations

New Obs	Length	Height
1	8.00	275

are two parameters in the simple linear regression model, so in multiple linear regression with p parameters, a logical estimator for σ^2 is

Estimator of Variance

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-p} = \frac{SS_E}{n-p} \tag{12.16}$$

This is an unbiased estimator of σ^2. Just as in simple linear regression, the estimate of σ^2 is usually obtained from the analysis of variance for the regression model. The numerator of Equation 12.16 is called the error or residual sum of squares, and the denominator $n - p$ is called the error or residual degrees of freedom.

We can find a computing formula for SS_E as follows:

$$SS_E = \sum_{i=1}^{n} (y_i - \hat{y})^2 = \sum_{i=1}^{n} e_i^2 = \mathbf{e}'\mathbf{e}$$

Substituting $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ into the equation, we obtain

$$SS_E = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}\mathbf{X}'\mathbf{y} = 27{,}178.5316 - 27{,}063.3581 = 115.174 \tag{12.17}$$

Table 12.4 shows that the estimate of σ^2 for the wire bond pull strength regression model is $\hat{\sigma}^2 = 115.2/22 = 5.2364$. The computer output rounds the estimate to $\hat{\sigma}^2 = 5.2$.

12.1.4 Properties of the Least Squares Estimators

The statistical properties of the least squares estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ may be easily found under certain assumptions on the error terms $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$, in the regression model. Paralleling the assumptions made in Chapter 11, we assume that the errors ϵ_i are statistically independent with mean zero and variance σ^2. Under these assumptions, the least squares estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ are unbiased estimators of the regression coefficients $\beta_0, \beta_1, \ldots, \beta_k$. This property may be shown as follows:

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon}] \\ &= \boldsymbol{\beta} \end{aligned}$$

because $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}$, the identity matrix. Thus, $\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$.

The variances of the $\hat{\boldsymbol{\beta}}$'s are expressed in terms of the elements of the inverse of the $\mathbf{X}'\mathbf{X}$ matrix. The inverse of $\mathbf{X}'\mathbf{X}$ times the constant σ^2 represents the **covariance matrix** of the regression coefficients $\hat{\boldsymbol{\beta}}$. The diagonal elements of $\sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$ are the variances of $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$, and the off-diagonal elements of this matrix are the covariances. For example, if we have $k = 2$ regressors, such as in the pull strength problem,

$$\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix}$$

which is symmetric ($C_{10} = C_{01}$, $C_{20} = C_{02}$, and $C_{21} = C_{12}$) because $(\mathbf{X}'\mathbf{X})^{-1}$ is symmetric, and we have

$$\begin{aligned} V(\hat{\beta}_j) &= \sigma^2 C_{jj}, \qquad j = 0, 1, 2 \\ \operatorname{cov}(\hat{\beta}_i, \hat{\beta}_j) &= \sigma^2 C_{ij}, \qquad i \neq j \end{aligned}$$

In general, the covariance matrix of $\hat{\boldsymbol{\beta}}$ is a $(p \times p)$ symmetric matrix whose jjth element is the variance of $\hat{\beta}_j$ and whose i, jth element is the covariance between $\hat{\beta}_i$ and $\hat{\beta}_j$, that is,

$$\operatorname{cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2\mathbf{C}$$

The estimates of the variances of these regression coefficients are obtained by replacing σ^2 with an estimate. When σ^2 is replaced by its estimate $\hat{\sigma}^2$, the square root of the estimated variance of the jth regression coefficient is called the **estimated standard error** of $\hat{\beta}_j$ or $se(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 C_{jj}}$. These standard errors are a useful measure of the **precision of estimation** for the regression coefficients; small standard errors imply good precision.

Multiple regression computer programs usually display these standard errors. For example, the computer output in Table 12.4 reports $se(\hat{\beta}_0) = 1.060$, $se(\hat{\beta}_1) = 0.09352$, and $se(\hat{\beta}_2) = 0.002798$. The intercept estimate is about twice the magnitude of its standard error, and $\hat{\beta}_1$ and β are considerably larger than $se(\hat{\beta}_1)$ and $se(\hat{\beta}_2)$. This implies reasonable precision of estimation, although the parameters β_1 and β_2 are much more precisely estimated than the intercept (this is not unusual in multiple regression).

12.2 Hypothesis Tests in Multiple Linear Regression

In multiple linear regression problems, certain tests of hypotheses about the model parameters are useful in measuring model adequacy. In this section, we describe several important hypothesis-testing procedures. As in the simple linear regression case, hypothesis testing requires that the error terms ϵ_i in the regression model are normally and independently distributed with mean zero and variance σ^2.

12.2.1 Test for Significance of Regression

The test for significance of regression is a test to determine whether a linear relationship exists between the response variable y and a subset of the regressor variables $x_1, x_2, \ldots, x_k$. The appropriate hypotheses are

Hypotheses for ANOVA Test

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

$$H_0: \beta_j \neq 0 \text{ for at least one } j \tag{12.18}$$

Rejection of $H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$ implies that at least one of the regressor variables $x_1, x_2, \ldots, x_k$ contributes significantly to the model.

The test for significance of regression is a generalization of the procedure used in simple linear regression. The total sum of squares SS_T is partitioned into a sum of squares due to the model or to regression and a sum of squares due to error, say,

$$SS_T = SS_R + SS_E$$

Now if $H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$ is true, SS_R/σ^2 is a chi-square random variable with k degrees of freedom. Note that the number of degrees of freedom for this chi-square random variable is equal to the number of regressor variables in the model. We can also show that the SS_E/σ^2 is a chi-square random variable with $n - p$ degrees of freedom, and that SS_E and SS_R are independent. The test statistic for $H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$ is

Test Statistic for ANOVA

$$F_0 = \frac{SS_R/k}{SS_E/(n-p)} = \frac{MS_R}{MS_E} \tag{12.19}$$

We should reject H_0 if the computed value of the test statistic in Equation 12.19, f_0, is greater than $f_{\alpha,k,n-p}$. The procedure is usually summarized in an analysis of variance table such as Table 12.5.

TABLE 12.5 **Analysis of Variance for Testing Significance of Regression in Multiple Regression**

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0
Regression	SS_R	k	MS_R	MS_R/MS_E
Error or residual	SS_E	$n-p$	MS_E	
Total	SS_T	$n-1$		

A computational formula for SS_R may be found easily. Now because $SS_T = \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2/n = \mathbf{y'y} - \left(\sum_{i=1}^{n} y_i\right)^2/n$, we may rewrite Equation 12.19 as

$$SS_E = \mathbf{y'y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} - \left[\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right]$$

or

$$SS_E = SS_T - SS_R \tag{12.20}$$

Therefore, the regression sum of squares is

$$SS_R = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \tag{12.21}$$

EXAMPLE 12.3 | Wire Bond Strength ANOVA

We test for significance of regression (with $\alpha = 0.05$) using the wire bond pull strength data from Example 12.1. The total sum of squares is

$$\begin{aligned} SS_T &= \mathbf{y'y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \\ &= 27{,}178.5316 - \frac{(725.82)^2}{25} \\ &= 6105.9447 \end{aligned}$$

The regression or model sum of squares is computed from Equation 12.21 as follows:

$$\begin{aligned} SS_R &= \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \\ &= 27{,}063.3581 - \frac{(725.82)^2}{25} = 5990.7712 \end{aligned}$$

and by subtraction

$$SS_E = SS_T - SS_R = \mathbf{y'y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = 115.1716$$

The analysis of variance is shown in Table 12.6. To test H_0: $\beta_1 = \beta_2 = 0$, we calculate the statistic

$$f_0 = \frac{MS_R}{MS_E} = \frac{2995.3856}{5.2352} = 572.17$$

Because $f_0 > f_{0.05,2,22} = 3.44$ (or because the P-value is considerably smaller than $\alpha = 0.05$), we reject the null hypothesis and conclude that pull strength is linearly related to either wire length or die height, or both.

Practical Interpretation: Rejection of H_0 does not necessarily imply that the relationship found is an appropriate model for predicting pull strength as a function of wire length and die height. Further tests of model adequacy are required before we can be comfortable using this model in practice.

TABLE 12.6 Test for Significance of Regression

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-value
Regression	5990.7712	2	2995.3856	572.17	1.08E-19
Error or residual	115.1735	22	5.2352		
Total	6105.9447	24			

Most multiple regression computer programs provide the test for significance of regression in their output display. The middle portion of Table 12.4 is the computer output for this example. Compare Tables 12.4 and 12.6 and note their equivalence apart from rounding. The P-value is rounded to zero in the computer output.

R^2 and Adjusted R^2 We may also use the **coefficient of multiple determination R^2** as a global statistic to assess the fit of the model. Computationally,

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{12.22}$$

For the wire bond pull strength data, we find that $R^2 = SS_R/SS_T = 5990.7712/6105.9447 = 0.9811$. Thus, the model accounts for about 98% of the variability in the pull strength response (refer to the computer software output in Table 12.4). The R^2 statistic is somewhat problematic as a measure of the quality of the fit for a multiple regression model because it never decreases when a variable is added to a model.

To illustrate, consider the model fit to the wire bond pull strength data in Example 11.8. This was a simple linear regression model with x_1= wire length as the regressor. The value of R^2 for this model is $R^2 = 0.9640$. Therefore, adding x_2 = die height to the model increases R^2 by $0.9811 - 0.9640 = 0.0171$, a very small amount. Because R^2 can never decrease when a regressor is added, it can be difficult to judge whether the increase is telling us anything useful about the new regressor. It is particularly hard to interpret a small increase, such as observed in the pull strength data.

Many regression users prefer to use an **adjusted** R^2 statistic:

Adjusted R^2

$$R^2_{\text{adj}} = 1 - \frac{SS_E/(n-p)}{SS_T/(n-1)} \tag{12.23}$$

Because $SS_E/(n-p)$ is the error or residual mean square and $SS_T/(n-p)$ is a constant, R^2_{adj} will only increase when a variable is added to the model if the new variable reduces the error mean square. Note that for the multiple regression model for the pull strength data $R^2_{\text{adj}} = 0.979$ (see the output in Table 12.4), whereas in Example 11.8, the adjusted R^2 for the one-variable model is $R^2_{\text{adj}} = 0.962$. Therefore, we would conclude that adding x_2 = die height to the model does result in a meaningful reduction in unexplained variability in the response.

The adjusted R^2 statistic essentially penalizes the analyst for adding terms to the model. It is an easy way to guard against **overfitting**, that is, including regressors that are not really

useful. Consequently, it is very useful in comparing and evaluating competing regression models.

12.2.2 Tests on Individual Regression Coefficients and Subsets of Coefficients

We are frequently interested in testing hypotheses on the individual regression coefficients. Such tests would be useful in determining the potential value of each of the regressor variables in the regression model. For example, the model might be more effective with the inclusion of additional variables or perhaps with the deletion of one or more of the regressors presently in the model.

The hypothesis to test if an individual regression coefficient, say β_j equals a value β_{j0} is

$$H_0: \beta_j = \beta_{j0} \qquad H_1: \beta_j \neq \beta_{j0} \tag{12.24}$$

The test statistic for this hypothesis is

$$T_0 = \frac{\hat{\beta}_j - \beta_{j0}}{\sqrt{\sigma^2 C_{jj}}} = \frac{\hat{\beta}_j - \beta_{j0}}{se(\hat{\beta}_j)} \tag{12.25}$$

where C_{jj} is the diagonal element of $(\mathbf{X'X})^{-1}$ corresponding to $\hat{\beta}_j$. Notice that the denominator of Equation 12.24 is the standard error of the regression coefficient $\hat{\beta}_j$. The null hypothesis H_0: $\beta_j = \beta_{j0}$ is rejected if $|t_0| > t_{\alpha/2,n-p}$. This is called a **partial** or **marginal test** because the regression coefficient $\hat{\beta}_j$ depends on all the other regressor variables $x_i (i \neq j)$ that are in the model. More will be said about this in the following example.

An important special case of the previous hypothesis occurs for $\beta_j = 0$. If H_0: $\beta_j = 0$ is not rejected, this indicates that the regressor x_j can be deleted from the model. Adding a variable to a regression model always causes the sum of squares for regression to increase and the error sum of squares to decrease (this is why R^2 always increases when a variable is added). We must decide whether the increase in the regression sum of squares is large enough to justify using the additional variable in the model. Furthermore, adding an unimportant variable to the model can actually increase the error mean square, indicating that adding such a variable has actually made the model a poorer fit to the data (this is why R^2_{adj} is a better measure of global model fit than the ordinary R^2).

EXAMPLE 12.4 | Wire Bond Strength Coefficient Test

Consider the wire bond pull strength data, and suppose that we want to test the hypothesis that the regression coefficient for x_2 (die height) is zero. The hypotheses are

$$H_0: \beta_2 = 0 \qquad H_1: \beta_2 \neq 0$$

The main diagonal element of the $(\mathbf{X'X})^{-1}$ matrix corresponding to $\hat{\beta}_2$ is $C_{22} = 0.0000015$, so the t-statistic in Equation 12.25 is

$$t_0 = \frac{\hat{\beta}_2}{\sqrt{\hat{\sigma}^2 C_{22}}} = \frac{0.01253}{\sqrt{(5.2352)(0.0000015)}} = 4.477$$

Note that we have used the estimate of σ^2 reported to four decimal places in Table 12.6. Because $t_{0.025,22} = 2.074$, we reject H_0: $\beta_2 = 0$ and conclude that the variable x_2 (die height) contributes significantly to the model. We could also have used a P-value to draw conclusions. The P-value for $t_0 = 4.477$ is $P = 0.0002$, so with $\alpha = 0.05$, we would reject the null hypothesis.

Practical Interpretation: Note that this test measures the marginal or partial contribution of x_2 given that x_1 is in the model. That is, the t-test measures the contribution of adding the variable x_2 = die height to a model that already contains x_1 = wire length. Table 12.4 shows the computer-generated value of the t-test computed. The computer software reports the t-test statistic to two decimal places. Note that the computer produces a t-test for each regression coefficient in the model. These t-tests indicate that both regressors contribute to the model.

EXAMPLE 12.5 | Wire Bond Strength One-Sided Coefficient Test

There is an interest in the effect of die height on strength. This can be evaluated by the magnitude of the coefficient for die height. To conclude that the coefficient for die height exceeds 0.01, the hypotheses become

$$H_0: \beta_2 = 0.01 \qquad H_1: \beta_2 > 0.01$$

For such a test, computer software can complete much of the hard work. We need only to assemble the pieces. From the output in Table 12.4, $\hat{\beta}_2 = 0.012528$, and the standard error of $\hat{\beta}_2 = 0.002798$. Therefore, the t-statistic is

$$t_0 = \frac{0.012528 - 0.01}{0.002798} = 0.9035$$

with 22 degrees of freedom (error degrees of freedom). From Table IV in Appendix A, $t_{0.25,22} = 0.686$ and $t_{0.1,22} = 1.321$. Therefore, the P-value can be bounded as $0.1 < P\text{-value} < 0.25$. One cannot conclude that the coefficient exceeds 0.01 at common levels of significance.

There is another way to test the contribution of an individual regressor variable to the model. This approach determines the increase in the regression sum of squares obtained by adding a variable x_j (say) to the model, given that other variables $x_i (i \neq j)$ are already included in the regression equation.

The procedure used to do this is called the **general regression significance test**, or the **extra sum of squares method**. This procedure can also be used to investigate the contribution of a *subset* of the regressor variables to the model. Consider the regression model with k regressor variables

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{12.26}$$

where $\mathbf{y}$ is $(n \times 1)$, $\mathbf{X}$ is $(n \times p)$, β is $(p \times 1)$, $\boldsymbol{\epsilon}$ is $(n \times 1)$, and $p = k + 1$. We would like to determine whether the subset of regressor variables $x_1, x_2, \ldots, x_r$ $(r < k)$ as a whole contributes significantly to the regression model. Let the vector of regression coefficients be partitioned as follows:

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} \tag{12.27}$$

where $\boldsymbol{\beta}_1$ is $(r \times 1)$ and $\boldsymbol{\beta}_2$ is $[(p - r) \times 1]$. We wish to test the hypotheses

Hypotheses for General Regression Test

$$H_0: \boldsymbol{\beta}_1 = \mathbf{0} \qquad H_1: \boldsymbol{\beta}_1 \neq \mathbf{0} \tag{12.28}$$

where $\mathbf{0}$ denotes a vector of zeroes. The model may be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon} \tag{12.29}$$

where $\mathbf{X}_1$ represents the columns of $\mathbf{X}$ associated with $\boldsymbol{\beta}_1$ and $\mathbf{X}_2$ represents the columns of $\mathbf{X}$ associated with $\boldsymbol{\beta}_2$.

For the **full model** (including both $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$), we know that $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. In addition, the regression sum of squares for all variables including the intercept is

$$SS_R(\boldsymbol{\beta}) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \qquad (p = k + 1 \text{ degrees of freedom})$$

and

$$MS_E = \frac{\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}\mathbf{X}'\mathbf{y}}{n - p}$$

$SS_R(\boldsymbol{\beta})$ is called the regression sum of squares due to $\boldsymbol{\beta}$. To find the contribution of the terms in $\boldsymbol{\beta}_1$ to the regression, fit the model assuming that the null hypothesis H_0: $\boldsymbol{\beta}_1 = \mathbf{0}$ to be true. The **reduced model** is found from Equation 12.29 as

$$\mathbf{y} = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon} \tag{12.30}$$

The least squares estimate of $\boldsymbol{\beta}_2$ is $\hat{\boldsymbol{\beta}}_2 = (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{y}$, and

$$SS_R(\boldsymbol{\beta}_2) = \hat{\boldsymbol{\beta}}_2\mathbf{X}_2'\mathrm{y} \qquad (p - \mathrm{r} \text{ degrees of freedom}) \tag{12.31}$$

The regression sum of squares due to $\boldsymbol{\beta}_1$ given that $\boldsymbol{\beta}_2$ is already in the model is

$$SS_R(\boldsymbol{\beta}_1|\boldsymbol{\beta}_2) = SS_R(\boldsymbol{\beta}) - SS_R(\boldsymbol{\beta}_2) \tag{12.32}$$

This sum of squares has r degrees of freedom. It is sometimes called the extra sum of squares due to $\boldsymbol{\beta}_1$. Note that $SS_R(\boldsymbol{\beta}_1|\boldsymbol{\beta}_2)$ is the increase in the regression sum of squares due to including the variables $x_1, x_2, \ldots, x_r$ in the model. Now $SS_R(\boldsymbol{\beta}_1|\boldsymbol{\beta}_2)$ is independent of MS_E, and the null hypothesis $\boldsymbol{\beta}_1 = \mathbf{0}$ may be tested by the statistic.

F Statistic for General Regression Tests

$$\mathrm{F}_0 = \frac{SS_R(\boldsymbol{\beta}_1|\boldsymbol{\beta}_2)/r}{MS_E} \tag{12.33}$$

If the computed value of the test statistic $f_0 > f_{\alpha,r,n-p}$, we reject H_0, concluding that at least one of the parameters in $\boldsymbol{\beta}_1$ is not zero and, consequently, at least one of the variables $x_1, x_2, \ldots, x_r$ in $\mathbf{X}_1$ contributes significantly to the regression model. Some authors call the test in Equation 12.33 a **partial F-test**.

The partial F-test is very useful. We can use it to measure the contribution of each individual regressor x_j as if it were the last variable added to the model by computing

$$SS_R(\beta_j|\beta_0, \beta_1, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_k), \qquad j = 1, 2, \ldots, k$$

This is the increase in the regression sum of squares due to adding x_j to a model that already includes $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k$. The partial F-test is a more general procedure in that we can measure the effect of sets of variables. The partial F-test plays a major role in *model building*—that is, in searching for the best set of regressor variables to use in the model.

EXAMPLE 12.6 | Wire Bond Strength General Regression Test

Consider the wire bond pull-strength data in Example 12.1. We investigate the contribution of two new variables, x_3 and x_4, to the model using the partial F-test approach. The new variables are explained at the end of this example. That is, we wish to test

$$H_0: \beta_3 = \beta_4 = 0 \qquad H_1: \beta_3 \neq 0 \text{ or } \beta_4 \neq 0$$

To test this hypothesis, we need the extra sum of squares due to β_3 and β_4 or

$$\begin{aligned} SS_R(\beta_4, \beta_3 | \beta_2, \beta_1, \beta_0) &= SS_R(\beta_4, \beta_3, \beta_2, \beta_1, \beta_0) - SS_R(\beta_2, \beta_1, \beta_0) \\ &= SS_R(\beta_4, \beta_3, \beta_2, \beta_1, |\beta_0) - SS_R(\beta_2, \beta_1, |\beta_0) \end{aligned}$$

In Example 12.3, we calculated

$$\begin{aligned} SS_R(\beta_2, \beta_1, |\beta_0) &= \mathbf{\beta' X' y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \\ &= 5990.7712 \quad \text{(2 degrees of freedom)} \end{aligned}$$

Also, Table 12.4 shows the computer output for the model with only x_1 and x_2 as predictors. In the analysis of variance table, we can see that $SS_R = 5990.8$, and this agrees with our calculation. In practice, the computer output would be used to obtain this sum of squares.

If we fit the model $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$, we can use the same matrix formula. Alternatively, we can look at SS_R from computer output for this model. The analysis of variance table for this model is shown in Table 12.7 and we see that

$$SS_R(\beta_4, \beta_3, \beta_2, \beta_1 | \beta_0) = 6024.0 \quad \text{(4 degrees of freedom)}$$

Therefore,

$$\begin{aligned} SS_R(\beta_4, \beta_3 | \beta_2, \beta_1, \beta_0) &= 6024.0 - 5990.8 \\ &= 33.2 \quad \text{(2 degrees of freedom)} \end{aligned}$$

This is the increase in the regression sum of squares due to adding x_3 and x_4 to a model already containing x_1 and x_2. To test H_0, calculate the test statistic

$$f_0 = \frac{SS_R(\beta_4, \beta_3 | \beta_2, \beta_1, \beta_0)/2}{MS_E} = \frac{33.2/2}{4.1} = 4.05$$

Note that MS_E from the full model using x_1, x_2, x_3, and x_4 is used in the denominator of the test statistic. Because $f_{0.05,2,20} = 3.49$, we reject H_0 and conclude that at least one of the new variables contributes significantly to the model. Further analysis and tests will be needed to refine the model and determine whether one or both of x_3 and x_4 are important.

TABLE 12.7 **Regression Analysis: *y* versus *x*1, *x*2, *x*3, *x*4**

The regression equation is y = 5.00 + 1.90 x1 + 0.0151 x2 + 0.0460 x3 − 0.000008 x4

Predictor	Coef	SE Coef	T	P
Constant	4.996	1.655	3.02	0.007
x1	1.9049	0.3126	6.09	0.000
x2	0.01513	0.01051	1.44	0.165
x3	0.04595	0.01666	2.76	0.012
x4	−0.00000766	0.00001641	−0.47	0.646
S = 2.02474	R − Sq = 98.75%	R − Sq(adj) = 98.4%		

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	6024.0	1506.0	367.35	0.000
Residual error	20	82.0	4.1		
Total	24	6105.9			

Source	DF	Seq SS
x1	1	5885.9
x2	1	104.9
x3	1	32.3
x4	1	0.9

The mystery of the new variables can now be explained. These are quadratic powers of the original predictors of wire length and wire height. That is, $x_3 = x_1^2$ and $x_4 = x_2^2$. A test for quadratic terms is a common use of partial F-tests. With this information and the original data for x_1 and x_2, we can use computer software to reproduce these calculations. Multiple regression allows models to be extended in such a simple manner that the real meaning of x_3 and x_4 did not even enter into the test procedure.

If a partial F-test is applied to a single variable, it is equivalent to a t-test. To see this, consider the computer software regression output for the wire bond pull strength in Table 12.4. Just below the analysis of variance summary in this table, the quantity labeled "'SeqSS"' shows the sum of squares obtained by fitting x_1 alone (5885.9) and the sum of squares obtained by fitting x_2 after x_1 (104.9). In our notation, these are referred to as $SS_R(\beta_1|\beta_0)$ and $SS_R(\beta_2, \beta_1|\beta_0)$, respectively. Therefore, to test H_0: $\beta_2 = 0$, H_1: $\beta_2 \neq 0$, the partial F-test is

$$f_0 = \frac{SS_R(\beta_2|\beta_1, \beta_0)/1}{MS_E} = \frac{104.92}{5.24} = 20.2$$

where MS_E is the mean square for residual in the computer output in Table 12.4. This statistic should be compared to an F-distribution with 1 and 22 degrees of freedom in the numerator and denominator, respectively. From Table 12.4, the t-test for the same hypothesis is $t_0 = 4.48$. Note that $t_0^2 = 4.48^2 = 20.07 = f_0$, except for round-off error. Furthermore, the square of a t-random variable with ν degrees of freedom is an F-random variable with 1 and ν degrees of freedom. Consequently, the t-test provides an equivalent method to test a single variable for contribution to a model. Because the t-test is typically provided by computer output, it is the preferred method to test a single variable.

12.3 Confidence Intervals in Multiple Linear Regression

12.3.1 Confidence Intervals on Individual Regression Coefficients

In multiple regression models, it is often useful to construct confidence interval estimates for the regression coefficients $\{\beta_j\}$. The development of a procedure for obtaining these confidence intervals requires that the errors $\{\epsilon_i\}$ are normally and independently distributed with mean zero and variance σ^2. This is the same assumption required in hypothesis testing. Therefore, the observations $\{Y_i\}$ are normally and independently distributed with mean $\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}$ and variance σ^2. Because the least squares estimator $\hat{\boldsymbol{\beta}}$ is a linear combination of the observations, it follows that $\hat{\boldsymbol{\beta}}$ is normally distributed with mean vector $\boldsymbol{\beta}$ and covariance matrix $\sigma^2(\mathbf{X'X})^{-1}$. Then each of the statistics

$$T = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \qquad j = 0, 1, \ldots, k \tag{12.34}$$

has a t distribution with $n - p$ degrees of freedom where C_{jj} is the jjth element of the $(\mathbf{X'X})^{-1}$ matrix, and $\hat{\sigma}^2$ is the estimate of the error variance, obtained from Equation 12.16. This leads to the following $100(1 - \alpha)\%$ confidence interval for the regression coefficient $\beta_j, j = 0, 1, \ldots, k$.

Confidence Interval on a Regression Coefficient

A $100(1-\alpha)\%$ **confidence interval on the regression coefficient** β_j, $j = 0, 1, \ldots, k$ in the multiple linear regression model is given by

$$\hat{\beta}_j - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 C_{jj}} \le \beta_j \le \hat{\beta}_j + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 C_{jj}} \tag{12.35}$$

Because $\sqrt{\hat{\sigma}^2 C_{jj}}$ is the standard error of the regression coefficient $\hat{\beta}_j$, we would also write the CI formula as $\hat{\beta}_j - t_{\alpha/2,n-p}\, se(\beta_j) \le \beta_j \le \hat{\beta}_j + t_{\alpha/2,n-p} se(\beta_j)$.

EXAMPLE 12.7 | Wire Bond Strength Confidence Interval

We construct a 95% confidence interval on the parameter β_1 in the wire bond pull strength problem. The point estimate of β_1 is $\hat{\beta}_1 = 2.74427$, and the diagonal element of $(\mathbf{X'X})^{-1}$ corresponding to β_1 is $C_{11} = 0.001671$. The estimate of σ^2 is $\hat{\sigma}^2 = 5.2352$ and $t_{0.025,22} = 2.074$. Therefore, the 95% CI on β_1 is computed from Equation 12.35 as

$$2.74427 - (2.074)\sqrt{(5.2352)(.001671)} \le \beta_1 \le 2.74427 + (2.074)\sqrt{(5.2352)(.001671)}$$

which reduces to

$$2.55029 \le \beta_1 \le 2.93825$$

Also, computer software can be used to help calculate this confidence interval. From the regression output in Table 10.4, $\hat{\beta}_1 = 2.74427$ and the standard error of $\hat{\beta}_1 = 0.0935$. This standard error is the multiplier of the t-table constant in the confidence interval. That is, $0.0935 = \sqrt{(5.2352)(0.001671)}$. Consequently, all the numbers are available from the computer output to construct the interval, which is the typical method used in practice.

12.3.2 Confidence Interval on the Mean Response

We may also obtain a confidence interval on the mean response at a particular point, say, $x_{01}, x_{02}, \ldots, x_{0k}$. To estimate the mean response at this point, define the vector

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ x_{01} \\ x_{02} \\ \vdots \\ x_{0k} \end{bmatrix}$$

The mean response at this point is $E(Y|\mathbf{x}_0) = \mu_{Y|\mathbf{x}_0} = \mathbf{x}_0'\boldsymbol{\beta}$, which is estimated by

$$\hat{\mu}_{Y|\mathbf{x}_0} = \mathbf{x}_0'\hat{\boldsymbol{\beta}} \tag{12.36}$$

This estimator is unbiased because $E(\mathbf{x}_0'\hat{\beta}) = \mathbf{x}_0'\boldsymbol{\beta} = E(Y|\mathbf{x}_0) = \mu_{Y|\mathbf{x}_0}$ and the variance of $\hat{\mu}_{Y|\mathbf{x}_0}$ is

$$V(\hat{\mu}_{Y|\mathbf{x}_0}) = \sigma^2 \mathbf{x}_0'(\mathbf{X'X})^{-1}\mathbf{x}_0 \tag{12.37}$$

A $100(1-\alpha)\%$ CI on $\mu_{Y|\mathbf{x}_0}$ can be constructed from the statistic

$$\frac{\hat{\mu}_{Y|\mathbf{x}_0} - \mu_{Y|\mathbf{x}_0}}{\sqrt{\hat{\sigma}^2 \mathbf{x}_0'(\mathbf{X'X})^{-1}\mathrm{x}_0}} \tag{12.38}$$

Confidence Interval on the Mean Response

For the multiple linear regression model, a $100(1-\alpha)\%$ **confidence interval on the mean response** at the point $x_{01}, x_{02}, \ldots, x_{0k}$ is

$$\hat{\mu}_{Y|\mathbf{x}_0} - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0} \le \mu_{Y|\mathbf{x}_0} \le \hat{\mu}_{Y|\mathbf{x}_0} + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0} \quad (12.39)$$

Equation 12.39 is a CI about the regression plane (or hyperplane). It is the multiple regression generalization in Equation 11.32.

EXAMPLE 12.8 | Wire Bond Strength Confidence Interval on the Mean Response

The engineer in Example 12.1 would like to construct a 95% CI on the mean pull strength for a wire bond with wire length $x_1 = 8$ and die height $x_2 = 275$. Therefore,

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 8 \\ 275 \end{bmatrix}$$

The estimated mean response at this point is found from Equation 12.36 as

$$\hat{\mu}_{Y|\mathbf{x}_0} = \mathbf{x}_0'\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1 & 8 & 275 \end{bmatrix}\begin{bmatrix} 2.26379 \\ 2.74427 \\ 0.01253 \end{bmatrix} = 27.66$$

The variance of $\hat{\mu}_{Y|\mathbf{x}_0}$ is estimated by

$$\hat{\sigma}^2\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 = 5.2352\begin{bmatrix} 1 & 8 & 275 \end{bmatrix}\begin{bmatrix} .214653 & -.007491 & -.000340 \\ -.007491 & .001671 & -.000019 \\ -.000340 & -.000019 & .0000015 \end{bmatrix}\begin{bmatrix} 1 \\ 8 \\ 275 \end{bmatrix}$$
$$= 5.2352(0.0444) = 0.23244$$

Therefore, a 95% CI on the mean pull strength at this point is found from Equation 12.39 as

$$27.66 - 2.074\sqrt{0.23244} \le \mu_{Y|\mathbf{x}_0} \le 27.66 + 2.074\sqrt{0.23244}$$

which reduces to

$$26.66 \le \mu_{Y|\mathbf{x}_0} \le 28.66$$

Some computer software will provide estimates of the mean for a point of interest $\mathbf{x_0}$ and the associated CI. Table 12.4 shows the computer output for Example 12.8. Both the estimate of the mean and the 95% CI are provided.

12.4 Prediction of New Observations

A regression model can be used to predict new or **future observations** on the response variable Y corresponding to particular values of the independent variables, say, $x_{01}, x_{02}, \ldots, x_{0k}$. If $\mathbf{x}_0' = [1, x_{01}, x_{02}, \ldots, x_{0k}]$, a point estimate of the future observation Y_0 at the point $x_{01}, x_{02}, \ldots, x_{0k}$ is

$$\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} \quad (12.40)$$

Prediction Interval

A $100(1-\alpha)\%$ **prediction interval** on a **future observation** is

$$\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)} \le Y_0 \le \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)} \quad (12.41)$$

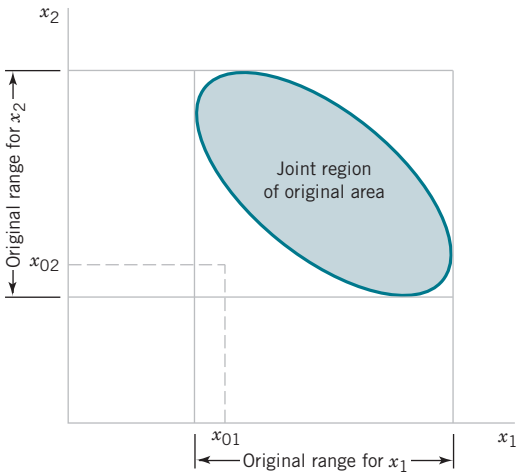


FIGURE 12.5

An example of extrapolation in multiple regression.

This prediction interval is a generalization of the prediction interval given in Equation 11.33 for a future observation in simple linear regression. If we compare the prediction interval Equation 12.41 with the expression for the confidence interval on the mean, Equation 12.39, you will observe that the prediction interval is always wider than the confidence interval. The confidence interval expresses the error in estimating the mean of a distribution, and the prediction interval expresses the error in predicting a future observation from the distribution at the point $\mathbf{x}_0$. This must include the error in estimating the mean at that point as well as the inherent variability in the random variable Y at the same value $\mathbf{x} = \mathbf{x}_0$.

Also, we might want to predict the mean of several values of Y, say m, all at the same value $\mathbf{x} = \mathbf{x}_0$. Because the variance of a sample mean is σ^2/m, Equation 12.41 is modified as follows. Replace the constant 1 under the square root with $1/m$ to reflect the lower variability in the mean of m observations. This results in a narrower interval.

In predicting new observations and in estimating the mean response at a given point $x_{01}, x_{02}, \ldots, x_{0k}$, we must be careful about **extrapolating** beyond the region containing the original observations. It is very possible that a model that fits well in the region of the original data will no longer fit well outside of that region. In multiple regression, inadvertently extrapolating is often easy because the levels of the variables $(x_{i1}, x_{i2}, \ldots, x_{ik})$, $i = 1, 2, \ldots, n$, jointly define the region containing the data. As an example, consider Figure 12.5, which illustrates the region containing the observations for a two-variable regression model. Note that the point (x_{01}, x_{02}) lies within the ranges of both regressor variables x_1 and x_2, but it is outside the region that is actually spanned by the original observations. This is sometimes called a **hidden extrapolation**. Either predicting the value of a new observation or estimating the mean response at this point is an extrapolation of the original regression model.

EXAMPLE 12.9 | Wire Bond Strength Confidence Interval

Suppose that the engineer in Example 12.1 wishes to construct a 95% prediction interval on the wire bond pull strength when the wire length is $x_1 = 8$ and the die height is $x_2 = 275$. Note that $\mathbf{x}_0' = [1 \;\; 8 \;\; 275]$, and the point estimate of the pull strength is $\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} = 27.66$. Also, in Example 12.8, we calculated $\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 = 0.04444$. Therefore, from Equation 12.41 we have

$$27.66 - 2.074\sqrt{5.2352(1 + 0.0444)} \leq Y_0 \leq 27.66 + 2.074\sqrt{5.2352(1 + 0.0444)}$$

and the 95% prediction interval is

$$22.81 \leq Y_0 \leq 32.51$$

Notice that the prediction interval is wider than the confidence interval on the mean response at the same point, calculated in Example 12.8. The computer output in Table 12.4 also displays this prediction interval.

12.5 Model Adequacy Checking

12.5.1 Residual Analysis

The **residuals** from the multiple regression model, defined by $e_i = y_i - \hat{y}_i$, play an important role in judging model adequacy just as they do in simple linear regression. As noted in Section 11.7.1, several residual plots are often useful; these are illustrated in Example 12.10. It is also helpful to plot the residuals against variables not presently in the model that are possible candidates for inclusion. Patterns in these plots may indicate that the model may be improved by adding the candidate variable.

EXAMPLE 12.10 | Wire Bond Strength Residuals

The residuals for the model from Example 12.1 are shown in Table 12.3. A normal probability plot of these residuals is shown in Figure 12.6. No severe deviations from normality are obviously apparent, although the two largest residuals ($e_{15} = 5.84$ and $e_{17} = 4.33$) do not fall extremely close to a straight line drawn through the remaining residuals.

The residuals are plotted against $\hat{y}$ in Figure 12.7, and against x_1 and x_2 in Figures 12.8 and 12.9, respectively.* The two largest residuals, e_{15} and e_{17}, are apparent. Figure 12.8 gives some indication that the model underpredicts the pull strength for assemblies with short wire length ($x_1 \leq 6$) and long wire length ($x_1 \geq 15$) and overpredicts the strength for assemblies with intermediate wire length ($7 \leq x_1 \leq 14$). The same impression is obtained from Figure 12.7. Either the relationship between strength and wire length is not linear (requiring that a term involving x_1^2, say, be added to the model) or other regressor variables not presently in the model affected the response.

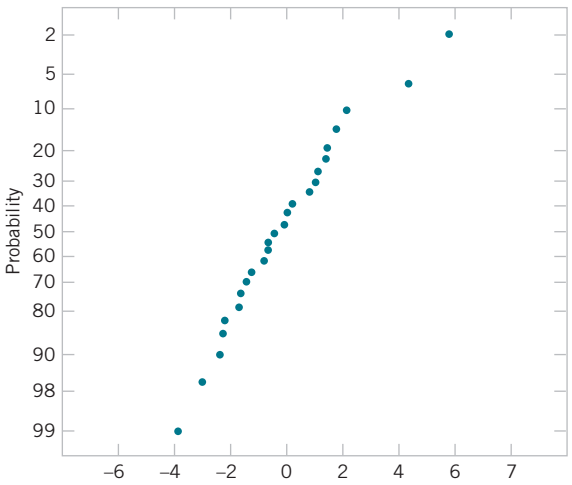


FIGURE 12.6
Normal probability plot of residuals.

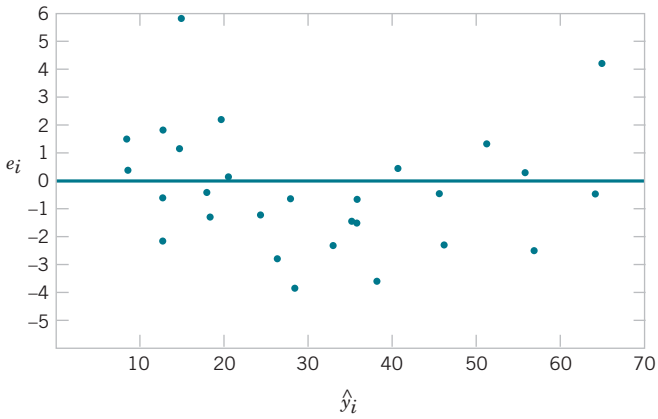


FIGURE 12.7
Plot of residuals against $\hat{y}$.

*There are other methods, such as those described in Montgomery, Peck, and Vining (2012) and Myers (1990), that plot a modified version of the residual, called a **partial residual**, against each regressor. These partial residual plots are useful in displaying the relationship between the response y and each individual regressor.

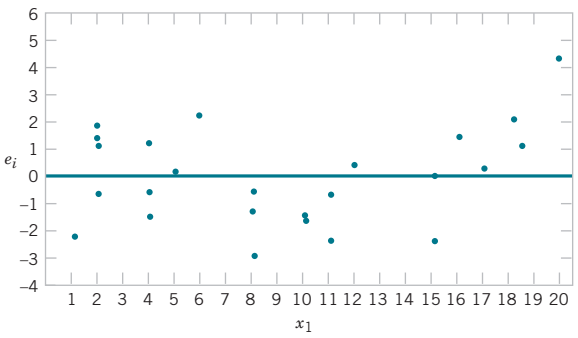


FIGURE 12.8
Plot of residuals against x_1.

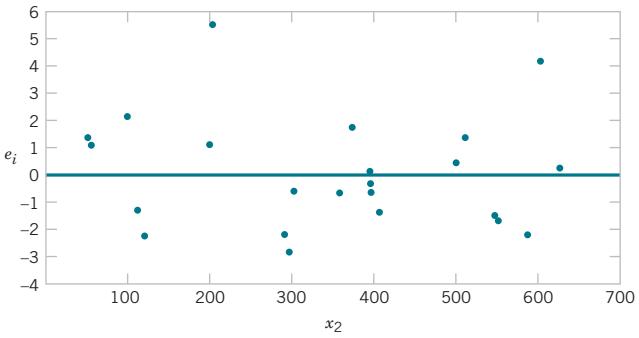


FIGURE 12.9
Plot of residuals against x_2.

The **standardized residuals**

Standardized Residual

$$d_i = \frac{e_i}{\sqrt{MS_E}} = \frac{e_i}{\sqrt{\hat{\sigma}^2}} \tag{12.42}$$

are often more useful than the ordinary residuals when assessing residual magnitude. For the wire bond strength example, the standardized residuals corresponding to e_{15} and e_{17} are $d_{15} = 5.84/\sqrt{5.2352} = 2.55$ and $d_{17} = 4.33/\sqrt{5.2352} = 1.89$, and they do not seem unusually large. Inspection of the data does not reveal any error in collecting observations 15 and 17, nor does it produce any other reason to discard or modify these two points.

In the wire bond strength example, we used the standardized residuals $d_i = e_i/\sqrt{\hat{\sigma}^2}$ as a measure of residual magnitude. Some analysts prefer to plot standardized residuals instead of ordinary residuals because the standardized residuals are scaled so that their standard deviation is approximately unity. Consequently, large residuals (that may indicate possible outliers or unusual observations) will be more obvious from inspection of the residual plots.

Many regression computer programs compute other types of scaled residuals. One of the most popular are the **studentized residuals**.

Studentized Residual

$$r_i = \frac{e_i}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}} \qquad i = 1, 2, \ldots, n \tag{12.43}$$

where h_{ii} is the ith diagonal element of the matrix

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

The $\mathbf{H}$ matrix is sometimes called the **hat matrix** because

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y}$$

Thus, $\mathbf{H}$ transforms the observed values of $\mathbf{y}$ into a vector of fitted values $\hat{\mathbf{y}}$.

Because each row of the matrix $\mathbf{X}$ corresponds to a vector, say $\mathbf{x}_i' = [1, x_{i1}, x_{i2}, \ldots, x_{ik}]$, another way to write the diagonal elements of the hat matrix is

Diagonal Elements of Hat Matrix

$$h_{ii} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i \tag{12.44}$$

Note that apart from σ^2, h_{ii} is the variance of the fitted value $\hat{y}_i$. The quantities h_{ii} were used in the computation of the confidence interval on the mean response in Section 12.3.2.

Under the usual assumptions that the model errors are independently distributed with mean zero and variance σ^2, we can show that the variance of the ith residual e_i is

$$V(e_i) = \sigma^2(1 - h_{ii}), \qquad i = 1, 2, \ldots, n$$

Furthermore, the h_{ii} elements must fall in the interval $0 < h_{ii} \leq 1$. This implies that the standardized residuals understate the true residual magnitude; thus, the studentized residuals would be a better statistic to examine in evaluating potential **outliers**.

To illustrate, consider the two observations identified in the wire bond strength data (Example 12.10) as having residuals that might be unusually large, observations 15 and 17. The standardized residuals are

$$d_{17} = \frac{e_{15}}{\sqrt{\hat{\sigma}^2}} = \frac{5.84}{\sqrt{5.2352}} = 2.55 \quad \text{and} \quad d_{17} = \frac{e_{17}}{\sqrt{MS_E}} = \frac{4.33}{\sqrt{5.2352}} = 1.89$$

Now $h_{15,15} = 0.0737$ and $h_{17,17} = 0.2593$, so the studentized residuals are

$$r_{15} = \frac{e_{15}}{\sqrt{\hat{\sigma}^2(1 - h_{15,15})}} = \frac{5.84}{\sqrt{5.2352(1 - 0.0737)}} = 2.65$$

and

$$r_{17} = \frac{e_{17}}{\sqrt{\hat{\sigma}^2(1 - h_{17,17})}} = \frac{4.33}{\sqrt{5.2352(1 - 0.2593)}} = 2.20$$

Notice that the studentized residuals are larger than the corresponding standardized residuals. However, the studentized residuals are still not so large as to cause us serious concern about possible outliers.

12.5.2 Influential Observations

When using multiple regression, we occasionally find that some subset of the observations is unusually influential. Sometimes these influential observations are relatively far away from the vicinity where the rest of the data were collected. A hypothetical situation for two variables is depicted in Figure 12.10 in which one observation in x-space is remote from the rest of the data. The disposition of points in the x-space is important in determining the properties of the model. For example, point (x_{i1}, x_{i2}) in Figure 12.10 may be very influential in determining R^2, the estimates of the regression coefficients, and the magnitude of the error mean square.

We would like to examine the influential points to determine whether they control many model properties. If these influential points are "bad" points, or erroneous in any way, they should be eliminated. On the other hand, there may be nothing wrong with these points, but at least we would like to determine whether or not they produce results consistent with the rest of the data. In any event, even if an influential point is a valid one, if it controls important model properties, we would like to know this, because it could have an impact on the use of the model.

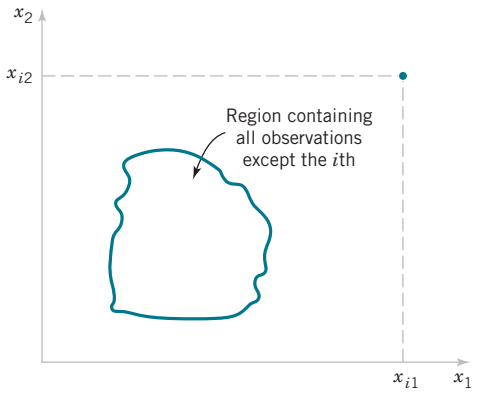


FIGURE 12.10
A point that is remote in x-space.

Montgomery, Peck, and Vining (2012) and Myers (1990) describe several methods for detecting influential observations. An excellent diagnostic is a measure of the squared distance between the usual least squares estimate of $\boldsymbol{\beta}$ based on all n observations and the estimate obtained when the ith point is removed, say, $\hat{\boldsymbol{\beta}}_{(i)}$. The **Cook's distance measure** is

Cook's Distance

$$D_i = \frac{(\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}})}{p\hat{\sigma}^2} \qquad i = 1, 2, \ldots, n$$

Clearly, if the ith point is influential, its removal will result in $\hat{\boldsymbol{\beta}}_{(i)}$ changing considerably from the value $\hat{\boldsymbol{\beta}}$. Thus, a large value of D_i implies that the ith point is influential. The statistic D_i is actually computed using

Cook's Distance Formula

$$D_i = \frac{r_i^2}{p}\frac{h_{ii}}{(1 - h_{ii})} \qquad i = 1, 2, \ldots, n \tag{12.45}$$

From Equation 12.45, we see that D_i consists of the squared studentized residual, which reflects how well the model fits the ith observation y_i [recall that $r_i = e_i/\sqrt{\hat{\sigma}^2(1 - h_{ii})}$] and a component that measures how far that point is from the rest of the data [$h_{ii}/(1 - h_{ii})$ is a measure of the distance of the ith point from the centroid of the remaining $n - 1$ points]. A value of $D_i > 1$ would indicate that the point is influential. Either component of D_i (or both) may contribute to a large value.

EXAMPLE 12.11 | Wire Bond Strength Cook's Distances

Table 12.8 lists the values of the hat matrix diagonals h_{ii} and Cook's distance measure D_i for the wire bond pull strength data in Example 12.1. To illustrate the calculations, consider the first observation:

$$D_1 = \frac{r_1^2}{p}\cdot\frac{h_{11}}{(1 - h_{11})} = -\frac{\left[e_1/\sqrt{MS_E(1 - h_{11})}\right]^2}{p}\cdot\frac{h_{11}}{(1 - h_{11})}$$
$$= \frac{\left[1.57/\sqrt{5.2352(1 - 0.1573)}\right]^2}{3}\cdot\frac{0.1573}{(1 - 0.1573)}$$
$$= 0.035$$

The Cook distance measure D_i does not identify any potentially influential observations in the data, for no value of D_i exceeds unity.

TABLE 12.8 Influence Diagnostics for the Wire Bond Pull Strength Data

Observations i	h_{ii}	Cook's Distance Measure D_i	Observations i	h_{ii}	Cook's Distance Measure D_i
1	0.1573	0.035	14	0.1129	0.003
2	0.1116	0.012	15	0.0737	0.187
3	0.1419	0.060	16	0.0879	0.001
4	0.1019	0.021	17	0.2593	0.565
5	0.0418	0.024	18	0.2929	0.155
6	0.0749	0.007	19	0.0962	0.018
7	0.1181	0.036	20	0.1473	0.000
8	0.1561	0.020	21	0.1296	0.052
9	0.1280	0.160	22	0.1358	0.028
10	0.0413	0.001	23	0.1824	0.002
11	0.0925	0.013	24	0.1091	0.040
12	0.0526	0.001	25	0.0729	0.000
13	0.0820	0.001			

12.6 Aspects of Multiple Regression Modeling

In this section, we briefly discuss several other aspects of building multiple regression models. These include polynomial terms, categorical variables, model building, and correlated regressors.

12.6.1 Polynomial Regression Models

The linear model $\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$ is a general model that can be used to fit any relationship that is **linear in the unknown parameters $\boldsymbol{\beta}$**. This includes the important class of **polynomial regression models**. For example, the second-degree polynomial in one variable

$$Y = \beta_0 + \beta_1 x + \beta_{11} x^2 + \varepsilon \tag{12.46}$$

and the second-degree polynomial in two variables

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon \tag{12.47}$$

are linear regression models.

Polynomial regression models are widely used when the response is curvilinear because the general principles of multiple regression can be applied. Example 12.12 illustrates some of the types of analyses that can be performed.

EXAMPLE 12.12 | Airplane Sidewall Panels

Sidewall panels for the interior of an airplane are formed in a 1500-ton press. The unit manufacturing cost varies with the production lot size. The following data give the average cost per unit (in hundreds of dollars) for this product (y) and the production lot size (x). The scatter diagram, shown in Figure 12.11, indicates that a second-order polynomial may be appropriate.

y	1.81	1.70	1.65	1.55	1.48	1.40
x	20	25	30	35	40	50

y	1.30	1.26	1.24	1.21	1.20	1.18
x	60	65	70	75	80	80

We fit the model

$$Y = \beta_0 + \beta_1 x + \beta_{11} x^2 + \varepsilon$$

The $\mathbf{y}$ vector, the model matrix $\mathbf{X}$, and the $\boldsymbol{\beta}$ vector are as follows:

$$\mathbf{y} = \begin{bmatrix} 1.81 \\ 1.70 \\ 1.65 \\ 1.55 \\ 1.48 \\ 1.40 \\ 1.30 \\ 1.26 \\ 1.24 \\ 1.21 \\ 1.20 \\ 1.18 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & 20 & 400 \\ 1 & 25 & 625 \\ 1 & 30 & 900 \\ 1 & 35 & 1225 \\ 1 & 40 & 1600 \\ 1 & 50 & 2500 \\ 1 & 60 & 3600 \\ 1 & 65 & 4225 \\ 1 & 70 & 4900 \\ 1 & 75 & 5625 \\ 1 & 80 & 6400 \\ 1 & 90 & 8100 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

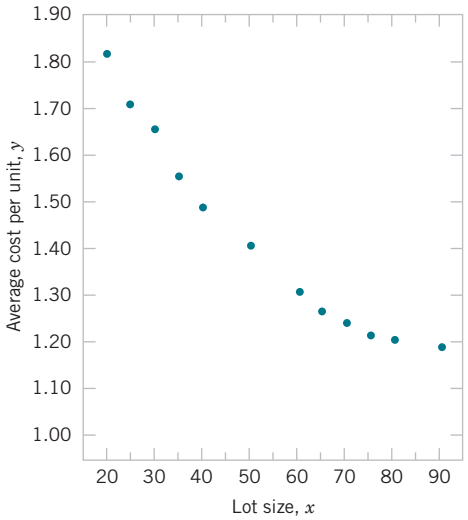


FIGURE 12.11 Data for Example 12.12

Solving the normal equations $\mathbf{X'X}\hat{\boldsymbol{\beta}} = \mathbf{X'y}$ gives the fitted model

$$\hat{y} = 2.19826629 - 0.02252236x + 0.00012507x^2$$

Conclusions: The test for significance of regression is shown in Table 12.9. Because $f_0 = 1762.3$ is significant at 1%, we conclude that at least one of the parameters β_1 and β_{11} is not zero. Furthermore, the standard tests for model adequacy do not reveal any unusual behavior, and we would conclude that this is a reasonable model for the sidewall panel cost data.

TABLE 12.9 Test for Significance of Regression for the Second-Order Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-value
Regression	0.52516	2	0.26258	1762.28	2.12E-12
Error	0.00134	9	0.00015		
Total	0.5265	11			

In fitting polynomials, we generally like to use the **lowest-degree model** consistent with the data. In this example, it would seem logical to investigate the possibility of dropping the quadratic term from the model. That is, we would like to test

$$H_0: \beta_{11} = 0 \quad H_1: \beta_{11} \neq 0$$

TABLE 12.10 Analysis of Variance for Example 12.12, Showing the Test for H_0: $\beta_{11} = 0$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-value
Regression	$SS_R(\beta_1, \beta_{11} \mid \beta_0) = 0.52516$	2	0.26258	1767.40	2.09E-12
Linear	$SS_R(\beta_1 \mid \beta_0) = 0.49416$	1	0.49416	2236.12	7.13E-13
Quadratic	$SS_R(\beta_{11} \mid \beta_0, \beta_1) = 0.03100$	1	0.03100	208.67	1.56E-7
Error	0.00133	9	0.00015		
Total	0.5265	11			

The general regression significance test can be used to test this hypothesis. We need to determine the "extra sum of squares" due to β_{11}, or

$$SS_R(\beta_{11} \mid \beta_1, \beta_0) = SS_R(\beta_1, \beta_{11} \mid \beta_0) - SS_R(\beta_1 \mid \beta_0)$$

The sum of squares $SS_R(\beta_1, \beta_{11}|\beta_0) = 0.52516$ from Table 12.10. To find $SS_R(\beta_1 \mid \beta_0)$, we fit a simple linear regression model to the original data, yielding

$$\hat{y} = 1.90036313 - 0.00910056x$$

It can be easily verified that the regression sum of squares for this model is

$$SS_R(\beta_1|\beta_0) = 0.4942$$

Therefore, the extra sum of the squares due to β_{11}, given that β_1 and β_0 are in the model, is

$$\begin{aligned} SS_R(\beta_{11}|\beta_1, \beta_0) &= SS_R(\beta_1, \beta_{11}|\beta_0) - SS_R(\beta_1|\beta_0) \\ &= 0.5252 - 0.4942 = 0.031 \end{aligned}$$

The analysis of variance with the test of H_0: $\beta_{11} = 0$ incorporated into the procedure is displayed in Table 12.10. Note that the quadratic term contributes significantly to the model.

12.6.2 Categorical Regressors and Indicator Variables

The regression models presented in previous sections have been based on **quantitative** variables, that is, variables that are measured on a numerical scale. For example, variables such as temperature, pressure, distance, and voltage are quantitative variables. Occasionally, we need to incorporate **categorical variables** in a regression model. For example, suppose that one of the variables in a regression model is the operator who is associated with each observation y_i. Assume that only two operators are involved. We may wish to assign different levels to the two operators to account for the possibility that each operator may have a different effect on the response.

The usual method of accounting for the different levels of a qualitative variable is to use **indicator variables**. For example, to introduce the effect of two different operators into a regression model, we could define an indicator variable as follows:

$$x = \begin{cases} 0 \text{ if the observation is from operator 1} \\ 1 \text{ if the observation is from operator 2} \end{cases}$$

In general, a qualitative variable with r-levels can be modeled by $r - 1$ indicator variables, which are assigned the value of either 0 or 1. Thus, if there are three operators, the different levels will be accounted for by the indicator variables defined as follows:

x_1	x_2	
0	0	if the observation is from operator 1
1	0	if the observation is from operator 2
0	1	if the observation is from operator 3

Indicator variables are also referred to as **dummy** variables. The following example [from Montgomery, Peck, and Vining (2012)] illustrates some of the uses of indicator variables.

EXAMPLE 12.13 | Surface Finish

A mechanical engineer is investigating the surface finish of metal parts produced on a lathe and its relationship to the speed (in revolutions per minute) of the lathe. The data are shown in Table 12.11. Note that the data have been collected using two different types of cutting tools. Because the type of cutting tool likely affects the surface finish, we fit the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

TABLE 12.11 Surface Finish Data

Observation Number, i	Surface Finish y_i	RPM	Type of Cutting Tool
1	45.44	225	302
2	42.03	200	302
3	50.10	250	302
4	48.75	245	302
5	47.92	235	302
6	47.79	237	302
7	52.26	265	302
8	50.52	259	302
9	45.58	221	302
10	44.78	218	302
11	33.50	224	416
12	31.23	212	416
13	37.52	248	416
14	37.13	260	416
15	34.70	243	416
16	33.92	238	416
17	32.13	224	416
18	35.47	251	416
19	33.49	232	416
20	32.29	216	416

where Y is the surface finish, x_1 is the lathe speed in revolutions per minute, and x_2 is an indicator variable denoting the type of cutting tool used; that is,

$$x_2 = \begin{cases} 0, \text{for tool type 302} \\ 1, \text{for tool type 416} \end{cases}$$

The parameters in this model may be easily interpreted. If $x_2 = 0$, the model becomes

$$Y = \beta_0 + \beta_1 x_1 + \varepsilon$$

which is a straight-line model with slope β_1 and intercept β_0. However, if $x_2 = 1$, the model becomes

$$Y = \beta_0 + \beta_1 x_1 + \beta_2(1) + \varepsilon = (\beta_0 + \beta_2) + \beta_1 x_1 + \varepsilon$$

which is a straight-line model with slope β_1 and intercept $\beta_0 + \beta_2$. Thus, the model $Y = \beta_0 + \beta_1 x + \beta_2 x_2 + \varepsilon$ implies that surface finish is linearly related to lathe speed and that the slope β_1 does not depend on the type of cutting tool used. However, the type of cutting tool does affect the intercept, and β_2 indicates the change in the intercept associated with a change in tool type from 302 to 416.

The model matrix $\mathbf{X}$ and $\mathbf{y}$ vector for this problem are as follows:

$$x = \begin{bmatrix} 1 & 225 & 1 \\ 1 & 200 & 1 \\ 1 & 250 & 1 \\ 1 & 245 & 1 \\ 1 & 235 & 1 \\ 1 & 237 & 1 \\ 1 & 265 & 1 \\ 1 & 259 & 1 \\ 1 & 221 & 1 \\ 1 & 218 & 1 \\ 1 & 224 & 1 \\ 1 & 212 & 1 \\ 1 & 248 & 1 \\ 1 & 260 & 1 \\ 1 & 243 & 1 \\ 1 & 238 & 1 \\ 1 & 224 & 1 \\ 1 & 251 & 1 \\ 1 & 232 & 1 \\ 1 & 216 & 1 \end{bmatrix} \quad y = \begin{bmatrix} 45.44 \\ 42.03 \\ 50.10 \\ 48.75 \\ 47.92 \\ 47.79 \\ 52.26 \\ 50.52 \\ 45.58 \\ 44.78 \\ 33.50 \\ 31.23 \\ 37.52 \\ 37.13 \\ 34.70 \\ 33.92 \\ 32.13 \\ 35.47 \\ 33.49 \\ 32.29 \end{bmatrix}$$

The fitted model is

$$\hat{y} = 14.27620 + 0.14115 x_1 - 13.28020 x_2$$

Conclusions: The analysis of variance for this model is shown in Table 12.12. Note that the hypothesis H_0: $\beta_1 = \beta_2 = 0$ (significance of regression) would be rejected at any reasonable level of significance because the P-value is very small. This table also contains the sums of squares

$$SS_R = SS_R(\beta_1, \beta_2|\beta_0) = SS_R(\beta_1|\beta_0| + SS_R(\beta_2)\beta_1, \beta_0)$$

so a test of the hypothesis H_0: $\beta_2 = 0$ can be made. Because this hypothesis is also rejected, we conclude that tool type has an effect on surface finish.

TABLE 12.12 Analysis of Variance

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-value
Regression	1012.0595	2	506.0297	1103.69	1.02E-18
$SS_R(\beta_1\|\beta_0)$	130.6091	1	130.6091	284.87	4.70E-12
$SS_R(\beta_2\|\beta_1, \beta_0)$	881.4504	1	881.4504	1922.52	6.24E-19
Error	7.7943	17	0.4585		
Total	1019.8538	19			

It is also possible to use indicator variables to investigate whether tool type affects both the slope and intercept. Let the model be

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

where x_2 is the indicator variable. Now if tool type 302 is used, $x_2 = 0$ and the model is

$$Y = \beta_0 + \beta_1 x_1 + \varepsilon$$

If tool type 416 is used, $x_2 = 1$ and the model becomes

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 + \beta_3 x_1 + \varepsilon = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)x_1 + \varepsilon$$

Note that β_2 is the change in the intercept and that β_3 is the change in slope produced by a change in tool type.

Another method of analyzing these data is to fit separate regression models to the data for each tool type. However, the indicator variable approach has several advantages. First, only one regression model must be fit. Second, by pooling the data on both tool types, more degrees of freedom for error are obtained. Third, tests of both hypotheses on the parameters β_2 and β_3 are just special cases of the extra sum of squares method.

12.6.3 Selection of Variables and Model Building

An important problem in many applications of regression analysis involves selecting the set of regressor variables to be used in the model. Sometimes previous experience or underlying theoretical considerations can help the analyst specify the set of regressor variables to use in a particular situation. Usually, however, the problem consists of selecting an appropriate set of regressors from a set that quite likely includes all the important variables, but we are sure that not all these candidate regressors are necessary to adequately model the response Y.

In such a situation, we are interested in **variable selection**; that is, screening the candidate variables to obtain a regression model that contains the "best" subset of regressor variables. We would like the final model to contain enough regressor variables so that in the intended use of the model (prediction, for example), it will perform satisfactorily. On the other hand, to keep

model maintenance costs to a minimum and to make the model easy to use, we would like the model to use as few regressor variables as possible. The compromise between these conflicting objectives is often called finding the "best" regression equation. However, in most problems, no single regression model is "best" in terms of the various evaluation criteria that have been proposed. A great deal of judgment and experience with the system being modeled is usually necessary to select an appropriate set of regressor variables for a regression equation.

No single algorithm will always produce a good solution to the variable selection problem. Most of the currently available procedures are search techniques, and to perform satisfactorily, they require interaction with judgment by the analyst. We now briefly discuss some of the more popular variable selection techniques. We assume that there are K candidate regressors, x_1, $x_2, \ldots, x_k$, and a single response variable y. All models will include an intercept term β_0, so the model with *all* variables included would have $K + 1$ terms. Furthermore, the functional form of each candidate variable (for example, $x_1 = 1/x$, $x_2 = \ln x$, etc.) is assumed to be correct.

All Possible Regressions This approach requires that the analyst fit all the regression equations involving one candidate variable, all regression equations involving two candidate variables, and so on. Then these equations are evaluated according to some suitable criteria to select the "best" regression model. If there are K candidate regressors, there are 2^K total equations to be examined. For example, if $K = 4$, there are $2^4 = 16$ possible regression equations; if $K = 10$, there are $2^{10} = 1024$ possible regression equations. Hence, the number of equations to be examined increases rapidly as the number of candidate variables increases. However, there are some very efficient computing algorithms for all possible regressions available and they are widely implemented in statistical software, so it is a very practical procedure unless the number of candidate regressors is fairly large. Look for a menu choice such as "Best Subsets" regression.

Several criteria may be used for evaluating and comparing the different regression models obtained. A commonly used criterion is based on the value of R^2 or the value of the adjusted R^2, R^2_{adj}. Basically, the analyst continues to increase the number of variables in the model until the increase in R^2 or the adjusted R^2_{adj} is small. Often, we will find that the R^2_{adj} will stabilize and actually begin to decrease as the number of variables in the model increases. Usually, the model that maximizes R^2_{adj} is considered to be a good candidate for the best regression equation. Because we can write $R^2_{\text{adj}} = 1 - \{MS_E/[SS_T/(n-1)]\}$ and $SS_T/(n-1)$ is a constant, the model that maximizes the R^2_{adj} value also minimizes the mean square error, so this is a very attractive criterion.

Another criterion used to evaluate regression models is the **C_p statistic**, which is a measure of the total mean square error for the regression model. We define the total standardized mean square error for the regression model as

$$\begin{aligned}\Gamma_p &= \frac{1}{\sigma^2}\sum_{i=1}^{n} E[\hat{Y}_i - E(Y_i)]^2 \\ &= \frac{1}{\sigma^2}\left\{\sum_{i=1}^{n}[E(Y_i) - E(\hat{Y}_i)]^2 + \sum_{i=1}^{n} V(\hat{Y}_i)\right\} \\ &= \frac{1}{\sigma^2}[(\text{bias})^2 + \text{variance}]\end{aligned}$$

We use the mean square error from the *full* $K + 1$ term model as an estimate of σ^2; that is, $\hat{\sigma}^2 = MS_E(K+1)$. Then an estimator of Γ_p is [see Montgomery, Peck, and Vining (2012) for the details]:

C_p Statistic

$$C_p = \frac{SS_E(p)}{\hat{\sigma}^2} - n + 2p \tag{12.48}$$

If the p-term model has negligible bias, it can be shown that

$$E(C_p|\text{zero bias}) = p$$

Therefore, the values of C_p for each regression model under consideration should be evaluated relative to p. The regression equations that have negligible bias will have values of C_p that are close to p, and those with significant bias will have values of C_p that are significantly greater than p. We then choose as the "best" regression equation either a model with *minimum* C_p or a model with a slightly larger C_p, that does not contain as much bias (i.e., $C_p \cong p$).

The **PRESS statistic** can also be used to evaluate competing regression models. PRESS is an acronym for **prediction error sum of squares**, and it is defined as the sum of the squares of the differences between each observation y_i and the corresponding predicted value based on a model fit to the *remaining* $n - 1$ points, say $\hat{y}_{(i)}$. So PRESS provides a measure of how well the model is likely to perform when predicting *new* data or data that were not used to fit the regression model. The computing formula for PRESS is

Prediction Error Sum of Squares (PRESS)

$$\text{PRESS} = \sum_{i=1}^{n}(y_i - \hat{y}_{(i)})^2 = \sum_{i=1}^{n}\left(\frac{e_i}{1 - h_{ii}}\right)^2$$

where $e_i = y_i - \hat{y}_i$ is the usual residual. Thus PRESS is easy to calculate from the standard least squares regression results. Models that have small values of PRESS are preferred.

EXAMPLE 12.14 | Wine Quality

Table 12.13 presents data on taste-testing 38 brands of pinot noir wine [the data were first reported in an article by Kwan, Kowalski, and Skogenboe in the *Journal Agricultural and Food Chemistry* (1979, Vol. 27), and it also appears as one of the default data sets in the Minitab software package]. The response variable is y = quality, and we wish to find the "best" regression equation that relates quality to the other five parameters.

Figure 12.12 is the matrix of scatter plots for the wine quality data. We notice that there are some indications of possible linear relationships between quality and the regressors,

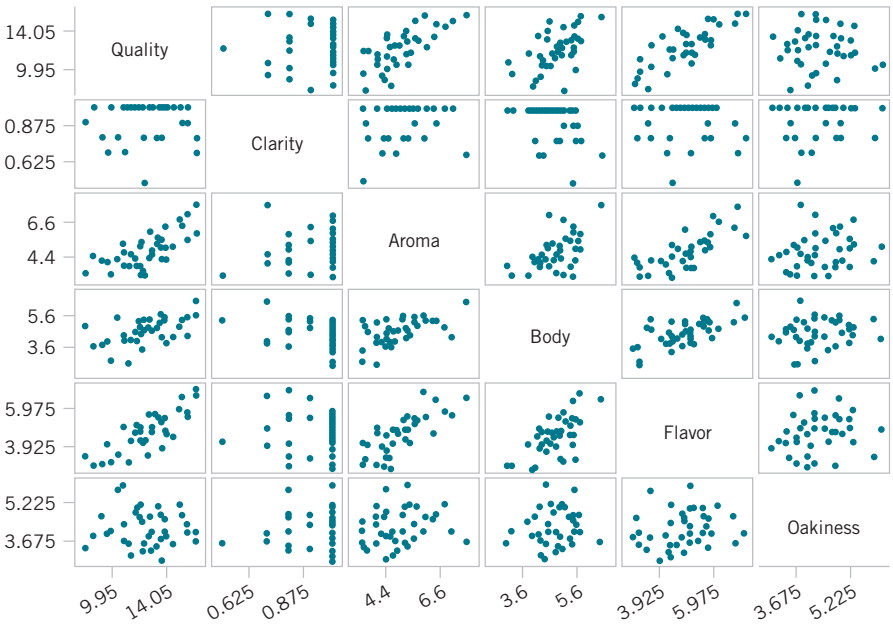


FIGURE 12.12 **A matrix of scatter plots from computer software for the wine quality data.**

TABLE 12.13 Wine Quality Data

	x_1 Clarity	x_2 Aroma	x_3 Body	x_4 Flavor	x_5 Oakiness	y Quality
1	1.0	3.3	2.8	3.1	4.1	9.8
2	1.0	4.4	4.9	3.5	3.9	12.6
3	1.0	3.9	5.3	4.8	4.7	11.9
4	1.0	3.9	2.6	3.1	3.6	11.1
5	1.0	5.6	5.1	5.5	5.1	13.3
6	1.0	4.6	4.7	5.0	4.1	12.8
7	1.0	4.8	4.8	4.8	3.3	12.8
8	1.0	5.3	4.5	4.3	5.2	12.0
9	1.0	4.3	4.3	3.9	2.9	13.6
10	1.0	4.3	3.9	4.7	3.9	13.9
11	1.0	5.1	4.3	4.5	3.6	14.4
12	0.5	3.3	5.4	4.3	3.6	12.3
13	0.8	5.9	5.7	7.0	4.1	16.1
14	0.7	7.7	6.6	6.7	3.7	16.1
15	1.0	7.1	4.4	5.8	4.1	15.5
16	0.9	5.5	5.6	5.6	4.4	15.5
17	1.0	6.3	5.4	4.8	4.6	13.8
18	1.0	5.0	5.5	5.5	4.1	13.8
19	1.0	4.6	4.1	4.3	3.1	11.3
20	.9	3.4	5.0	3.4	3.4	7.9
21	0.9	6.4	5.4	6.6	4.8	15.1
22	1.0	5.5	5.3	5.3	3.8	13.5
23	0.7	4.7	4.1	5.0	3.7	10.8
24	0.7	4.1	4.0	4.1	4.0	9.5
25	1.0	6.0	5.4	5.7	4.7	12.7
26	1.0	4.3	4.6	4.7	4.9	11.6
27	1.0	3.9	4.0	5.1	5.1	11.7
28	1.0	5.1	4.9	5.0	5.1	11.9
29	1.0	3.9	4.4	5.0	4.4	10.8
30	1.0	4.5	3.7	2.9	3.9	8.5
31	1.0	5.2	4.3	5.0	6.0	10.7
32	0.8	4.2	3.8	3.0	4.7	9.1
33	1.0	3.3	3.5	4.3	4.5	12.1
34	1.0	6.8	5.0	6.0	5.2	14.9
35	0.8	5.0	5.7	5.5	4.8	13.5
36	0.8	3.5	4.7	4.2	3.3	12.2
37	0.8	4.3	5.5	3.5	5.8	10.3
38	0.8	5.2	4.8	5.7	3.5	13.2

but there is no obvious visual impression of which regressors would be appropriate. Table 12.14 lists all possible regressions output from the software. In this analysis, we asked the computer software to present the best three equations for each subset size. Note that the computer software reports the values of R^2, R^2_{adj}, C_p, and $S = \sqrt{MS_E}$ for each model. From Table 12.14 we see that the three-variable equation with x_2 = aroma, x_4 = flavor, and x_5 = oakiness produces the minimum C_p equation, whereas the four-variable model, which adds x_1 = clarity to the previous three regressors, results in maximum R^2_{adj} (or minimum MS_E). The three-variable model is

$$\hat{y} = 6.47 + 0.580x_2 + 1.20x_4 - 0.602x_5$$

and the four-variable model is

$$\hat{y} = 4.99 + 1.79\,x_1 + 0.530\,x_2 + 1.26\,x_4 - 0.659\,x_5$$

These models should now be evaluated further using residual plots and the other techniques discussed earlier in the chapter to see whether either model is satisfactory with respect to the underlying assumptions and to determine whether one of them is preferable. It turns out that the residual plots do not reveal any major problems with either model. The value of PRESS for the three-variable model is 56.0524, and for the four-variable model, it is 60.3327. Because PRESS is smaller in the model with three regressors and because it is the model with the smallest number of predictors, it would likely be the preferred choice.

TABLE 12.14 All Possible Regressions Computer Output for the Wine Quality Data

```
Best Subsets Regression: Quality versus Clarity, Aroma, …
Response is quality
```

Vars	R-Sq	R-Sq (adj)	C-p	S	Clarity	Aroma	Body	Flavor	Oakiness
1	62.4	61.4	9.0	1.2712				X	
1	50.0	48.6	23.2	1.4658		X			
1	30.1	28.2	46.0	1.7335			X		
2	66.1	64.2	6.8	1.2242				X	X
2	65.9	63.9	7.1	1.2288		X		X	
2	63.3	61.2	10.0	1.2733	X			X	
3	70.4	67.8	3.9	1.1613		X		X	X
3	68.0	65.2	6.6	1.2068	X			X	X
3	66.5	63.5	8.4	1.2357			X	X	X
4	71.5	68.0	4.7	1.1568	X	X		X	X
4	70.5	66.9	5.8	1.1769		X	X	X	X
4	69.3	65.6	7.1	1.1996	X		X	X	X
5	72.1	67.7	6.0	1.1625	X	X	X	X	X

Stepwise Regression **Stepwise regression** is probably the most widely used variable selection technique. The procedure iteratively constructs a sequence of regression models by adding or removing variables at each step. The criterion for adding or removing a variable at any step is

usually expressed in terms of a partial F-test. Let f_{in} be the value of the F-random variable for adding a variable to the model, and let f_{out} be the value of the F-random variable for removing a variable from the model. We must have $f_{\text{in}} \geq f_{\text{out}}$, and usually $f_{\text{in}} = f_{\text{out}}$.

Stepwise regression begins by forming a one-variable model using the regressor variable that has the highest correlation with the response variable Y. This will also be the regressor producing the largest F-statistic. For example, suppose that at this step, x_1 is selected. At the second step, the remaining $K - 1$ candidate variables are examined, and the variable for which the partial F-statistic

$$F_j = \frac{SS_R(\beta_j|\beta_1, \beta_0)}{MS_E(x_j, x_1)} \tag{12.49}$$

is a maximum is added to the equation provided that $f_j > f_{\text{in}}$. In Equation 12.4, $MS_E(x_j, x_1)$ denotes the mean square for error for the model containing both x_1 and x_j. Suppose that this procedure indicates that x_2 should be added to the model. Now the stepwise regression algorithm determines whether the variable x_1 added at the first step should be removed. This is done by calculating the F-statistic

$$F_1 = \frac{SS_R(\beta_1|\beta_2, \beta_0)}{MS_E(x_1, x_2)} \tag{12.50}$$

If the calculated value $f_1 < f_{\text{out}}$, the variable x_1 is removed; otherwise it is retained, and we would attempt to add a regressor to the model containing both x_1 and x_2.

In general, at each step the set of remaining candidate regressors is examined, and the regressor with the largest partial F-statistic is entered provided that the observed value of f exceeds f_{in}. Then the partial F-statistic for each regressor in the model is calculated, and the regressor with the smallest observed value of F is deleted if the observed $f < f_{\text{out}}$. The procedure continues until no other regressors can be added to or removed from the model.

Stepwise regression is almost always performed using a computer program. The analyst exercises control over the procedure by the choice of f_{in} and f_{out}. Some stepwise regression computer programs require that numerical values be specified for f_{in} and f_{out}. Because the number of degrees of freedom on MS_E depends on the number of variables in the model, which changes from step to step, a fixed value of f_{in} and f_{out} causes the type I and type II error rates to vary. Some computer programs allow the analyst to specify the type I error levels for f_{in} and f_{out}. However, the "advertised" significance level is not the true level because the variable selected is the one that maximizes (or minimizes) the partial F-statistic at that stage. Sometimes it is useful to experiment with different values of f_{in} and f_{out} (or different advertised type I error rates) in several different runs to see whether this substantially affects the choice of the final model.

EXAMPLE 12.15 | Wine Quality Stepwise Regression

Table 12.15 gives the software stepwise regression output for the wine quality data. The software uses fixed values of a for entering and removing variables. The default level is $\alpha = 0.15$ for both decisions. The output in Table 12.12 uses the default value. Notice that the variables were entered in the order flavor (step 1), oakiness (step 2), and aroma (step 3) and that no variables were removed. No other variable could be entered, so the algorithm terminated. This is the three-variable model found by all possible regressions that results in a minimum value of C_p.

TABLE 12.15 **Stepwise Regression Output for the Wine Quality Data**

```
Stepwise Regression: Quality versus Clarity, Aroma, ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is Quality on 5 predictors, with N = 38

Step                 1          2          3
Constant         4.941      6.912      6.467
Flavor            1.57       1.64       1.20
T-Value           7.73       8.25       4.36
P-Value          0.000      0.000      0.000

Oakiness                   −0.54      −0.60
T-Value                    −1.95      −2.28
P-Value                    0.059      0.029

Aroma                                 0.58
T-Value                               2.21
P-Value                               0.034

S                 1.27       1.22       1.16
R-Sq             62.42      66.11      70.38
R-Sq(adj)        61.37      64.17      67.76
C-p                9.0        6.8        3.9
```

Forward Selection The **forward selection** procedure is a variation of stepwise regression and is based on the principle that regressors should be added to the model one at a time until there are no remaining candidate regressors that produce a significant increase in the regression sum of squares. That is, variables are added one at a time as long as their partial F-value exceeds f_{in}. Forward selection is a simplification of stepwise regression that omits the partial F-test for deleting variables from the model that have been added at previous steps. This is a potential weakness of forward selection; that is, the procedure does not explore the effect that adding a regressor at the current step has on regressor variables added at earlier steps. Notice that if we were to apply forward selection to the wine quality data, we would obtain exactly the same results as we did with stepwise regression in Example 12.15 because stepwise regression terminated without deleting a variable.

Backward Elimination The **backward elimination** algorithm begins with all K candidate regressors in the model. Then the regressor with the smallest partial F-statistic is deleted if this F-statistic is insignificant, that is, if $f < f_{\text{out}}$. Next, the model with $K - 1$ regressors is fit, and the next regressor for potential elimination is found. The algorithm terminates when no further regressor can be deleted.

TABLE 12.16 Backward Elimination Output for the Wine Quality Data

Stepwise Regression: Quality versus Clarity, Aroma, ...

Backward elimination. Alpha-to-Remove: 0.1

Response is Quality on 5 predictors, with N = 38

Step	1	2	3
Constant	3.997	4.986	6.467
Clarity	2.3	1.8	
T-Value	1.35	1.12	
P-Value	0.187	0.269	
Aroma	0.48	0.53	0.58
T-Value	1.77	2.00	2.21
P-Value	0.086	0.054	0.034
Body	0.27		
T-Value	0.82		
P-Value	0.418		
Flavor	1.17	1.26	1.20
T-Value	3.84	4.52	4.36
P-Value	0.001	0.000	0.000
Oakiness	−0.68	−0.66	−0.60
T-Value	−2.52	−2.46	−2.28
P-Value	0.017	0.019	0.029
S	1.16	1.16	1.16
R-Sq	72.06	71.47	70.38
R-Sq(adj)	67.69	68.01	67.76
C-p	6.0	4.7	3.9

Table 12.16 shows the computer software package output for backward elimination applied to the wine quality data. The α value for removing a variable is $\alpha = 0.10$. Notice that this procedure removes body at step 1 and then clarity at step 2, terminating with the three-variable model found previously.

Some Comments on Final Model Selection We have illustrated several different approaches to the selection of variables in multiple linear regression. The final model obtained from any model-building procedure should be subjected to the usual adequacy checks, such as residual analysis, lack-of-fit testing, and examination of the effects of influential points. The analyst may also consider augmenting the original set of candidate variables with cross-products, polynomial terms, or other transformations of the original variables that might improve the model. A major criticism of variable selection methods such as stepwise regression is that the analyst may conclude that there is one "best" regression equation. Generally, this is not the case because several equally good regression models can often be used. One way to avoid this problem is to use

several different model-building techniques and see whether different models result. For example, we have found the same model for the wine quality data using stepwise regression, forward selection, and backward elimination. The same model was also one of the two best found from all possible regressions. The results from variable selection methods frequently do not agree, so this is a good indication that the three-variable model is the best regression equation.

If the number of candidate regressors is not too large, the all-possible regressions method is recommended. We usually recommend using the minimum MS_E and C_p evaluation criteria in conjunction with this procedure. The all-possible regressions approach can find the "best" regression equation with respect to these criteria, but stepwise-type methods offer no such assurance. Furthermore, the all-possible regressions procedure is not distorted by dependencies among the regressors as stepwise-type methods are.

12.6.4 Multicollinearity

In multiple regression problems, we expect to find dependencies between the response variable Y and the regressors x_j. In most regression problems, however, we find that there are also dependencies among the regressor variables x_j. In situations in which these dependencies are strong, we say that **multicollinearity** exists. Multicollinearity can have serious effects on the estimates of the regression coefficients and on the general applicability of the estimated model.

The effects of multicollinearity may be easily demonstrated. The diagonal elements of the matrix $C = (\mathbf{X'X})^{-1}$ can be written as

$$C_{jj} = \frac{1}{(1 - R_j^2)} \quad j = 1,2,\ldots,k$$

where R_j^2 is the coefficient of multiple determination resulting from regressing x_j on the other $k - 1$ regressor variables. We can think of R_j^2 as a measure of the correlation between x_j and the other regressors. Clearly, the stronger the linear dependency of x_j on the remaining regressor variables and hence the stronger the multicollinearity, the greater the value of R_j^2 will be. Recall that $V(\hat{\beta}_j) = \sigma^2 C_{jj}$. Therefore, we say that the variance of $\hat{\beta}_j$ is "inflated" by the quantity $(1 - R_j^2)^{-1}$. Consequently, we define the **variance inflation factor** for β_j as

Variance Inflation Factor (VIF)

$$VIF(\beta_j) = \frac{1}{(1 - R_j^2)} \quad j = 1,2,\ldots,k \tag{12.51}$$

These factors are important measures of the extent to which multicollinearity is present. If the columns of the model matrix $\mathbf{X}$ are **orthogonal**, then the regressors are completely uncorrelated, and the variance inflation factors will all be unity. So, any VIF that exceeds 1 indicates some level of multicollinearity in the data.

Although the estimates of the regression coefficients are very imprecise when multicollinearity is present, the fitted model equation may still be useful. For example, suppose that we wish to predict new observations on the response. If these predictions are interpolations in the original region of the x-space where the multicollinearity is in effect, satisfactory predictions will often be obtained because while individual β_j may be poorly estimated, the function $\sum_{j=1}^{k} \beta_j x_{ij}$ may be estimated quite well. On the other hand, if the prediction of new observations requires extrapolation beyond the original region of the x-space where the data were collected, generally we would expect to obtain poor results. Extrapolation usually requires good estimates of the individual model parameters.

Multicollinearity arises for several reasons. It will occur when the analyst collects data such that a linear constraint holds approximately among the columns of the $\mathbf{X}$ matrix. For example,

if four regressor variables are the components of a mixture, such a constraint will always exist because the sum of the components is always constant. Usually, these constraints do not hold exactly, and the analyst might not know that they exist.

The presence of multicollinearity can be detected in several ways. Two of the more easily understood of these are discussed briefly here.

1. The variance inflation factors, defined in Equation 12.5, are very useful measures of multicollinearity. The larger the variance inflation factor, the more severe the multicollinearity. Some authors have suggested that if any variance inflation factor exceeds 10, multicollinearity is a problem. Other authors consider this value too liberal and suggest that the variance inflation factors should not exceed 4 or 5. Computer software will calculate the variance inflation factors. Table 12-4 presents the computer-generated multiple regression output for the wire bond pull strength data. Because both VIF_1 and VIF_2 are small, there is no problem with multicollinearity.
2. If the F-test for significance of regression is significant but tests on the individual regression coefficients are not significant, multicollinearity may be present.

Several remedial measures have been proposed for solving the problem of multicollinearity. Augmenting the data with new observations specifically designed to break up the approximate linear dependencies that currently exist is often suggested. However, this is sometimes impossible because of economic reasons or because of the physical constraints that relate to x_j. Another possibility is to delete certain variables from the model, but this approach has the disadvantage of discarding the information contained in the deleted variables.

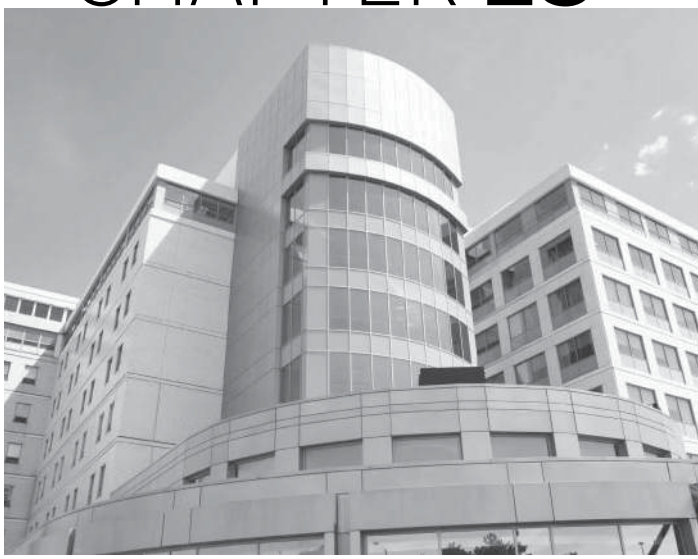
Because multicollinearity primarily affects the stability of the regression coefficients, it would seem that estimating these parameters by some method that is less sensitive to multicollinearity than ordinary least squares would be helpful. Several methods have been suggested. One alternative to ordinary least squares, **ridge regression**, can be useful in combating multicollinearity. For more details on ridge regression, there are more extensive presentations in Montgomery, Peck, and Vining (2012).

Important Terms and Concepts

All possible regressions
Analysis of variance test in multiple regression
Backward elimination
C_p statistic
Categorical variables
Confidence interval on the mean response
Confidence interval on the regression coefficient
Cook's distance measure
Extra sum of squares method
Forward selection
Full model
Hat matrix
Hidden extrapolation
Indicator variables
Inference (tests and intervals) on individual model parameters
Influential observations
Model parameters and their interpretation in multiple regression
Multicollinearity
Multiple regression model
Outliers
Partial or marginal test
Polynomial regression model
Prediction interval on a future observation
PRESS statistic
R^2
Residual analysis and model adequacy checking
Significance of regression
Standardized residuals
Stepwise regression
Studentized residuals
Variable selection
Variance inflation factor (VIF)

CHAPTER **13**

Design and Analysis of Single-Factor Experiments: The Analysis of Variance



© Vasko Miokovic/iStockphoto

LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Design and conduct engineering experiments involving a single factor with an arbitrary number of levels
2. Understand how the analysis of variance is used to analyze the data from these experiments
3. Assess model adequacy with residual plots
4. Use multiple comparison procedures to identify specific differences between means
5. Make decisions about sample size in single-factor experiments
6. Understand the difference between fixed and random factors
7. Estimate variance components in an experiment with random factors
8. Understand the blocking principle and how it is used to isolate the effect of nuisance factors
9. Design and analyze experiments involving the randomized complete block design

CHAPTER OUTLINE

13.1 Designing Engineering Experiments

13.2 Completely Randomized Single-Factor Experiment

- **13.2.1** Example: Tensile Strength
- **13.2.2** Analysis of Variance
- **13.2.3** Multiple Comparisons Following the ANOVA
- **13.2.4** Residual Analysis and Model Checking
- **13.2.5** Determining Sample Size

13.3 The Random-Effects Model

- **13.3.1** Fixed Versus Random Factors
- **13.3.2** ANOVA and Variance Components

13.4 Randomized Complete Block Design

- **13.4.1** Design and Statistical Analysis
- **13.4.2** Multiple Comparisons
- **13.4.3** Residual Analysis and Model Checking

Experiments are a natural part of the engineering and scientific decision-making process. Suppose, for example, that a civil engineer is investigating the effects of different curing methods on the mean compressive strength of concrete. The experiment would consist of making up several test specimens of concrete using each of the proposed curing methods and then testing the compressive strength of each specimen. The data from this experiment could be used to determine which curing method should be used to provide maximum mean compressive strength.

If there are only two curing methods of interest, this experiment could be designed and analyzed using the statistical hypothesis methods for two samples introduced in Chapter 10. That is, the experimenter has a single **factor** of interest—curing methods—and there are only two **levels** of the factor. If the experimenter is interested in determining which curing method produces the maximum compressive strength, the number of specimens to test can be determined from the operating characteristic curves in Appendix Chart VII, and the t-test can be used to decide if the two means differ.

Many single-factor experiments require that more than two levels of the factor be considered. For example, the civil engineer may want to investigate five different curing methods. In this chapter, we show how the **analysis of variance** (frequently abbreviated ANOVA) can be used for comparing means when there are more than two levels of a single factor. We also discuss randomization of the experimental runs and the important role this concept plays in the overall experimentation strategy. In the next chapter, we show how to design and analyze experiments with several factors.

13.1 Designing Engineering Experiments

Statistically based experimental design techniques are particularly useful in the engineering world for solving many important problems: discovery of new basic phenomena that can lead to new products and commercialization of new technology including new product development, new process development, and improvement of existing products and processes. For example, consider the development of a new process. Most processes can be described in terms of several controllable variables, such as temperature, pressure, and feed rate. By using designed experiments, engineers can determine which subset of the process variables has the greatest influence on process performance. The results of such an experiment can lead to:

- Improved process yield
- Reduced variability in the process and closer conformance to nominal or target requirements
- Reduced design and development time
- Reduced cost of operation

Experimental design methods are also useful in engineering design activities during which new products are developed and existing ones are improved. Some typical applications of statistically designed experiments in engineering design include:

- Evaluation and comparison of basic design configurations
- Evaluation of different materials
- Selection of design parameters so that the product will work well under a wide variety of field conditions (or so that the design will be robust)
- Determination of key product design parameters that affect product performance

The use of experimental design in the engineering design process can result in products that are easier to manufacture, products that have better field performance and reliability than their competitors, and products that can be designed, developed, and produced in less time.

Designed experiments are usually employed *sequentially*. That is, the first experiment with a complex system (perhaps a manufacturing process) that has many controllable variables is often a *screening experiment* designed to determine those variables that are most important. Subsequent experiments are used to refine this information and determine which adjustments to these critical variables are required to improve the process. Finally, the objective of the experimenter is *optimization*, that is, to determine those levels of the critical variables that result in the best process performance.

The statistical methods introduced in this chapter and Chapter 14 are essential to good experimentation. *All experiments are designed experiments*; unfortunately, some of them are poorly designed, and, as a result, valuable resources are used ineffectively. Statistically designed experiments permit efficiency and economy in the experimental process, and the use of statistical methods in examining the data results in *scientific objectivity* when drawing conclusions.

13.2 Completely Randomized Single-Factor Experiment

13.2.1 Example: Tensile Strength

A manufacturer of paper is interested in improving the product's tensile strength. Product engineering believes that tensile strength is a function of the hardwood concentration in the pulp and that the range of hardwood concentrations of practical interest is between 5 and 20%. A team of engineers responsible for the study decides to investigate four levels of hardwood concentration: 5%, 10%, 15%, and 20%. They decide to make up six test specimens at each concentration level by using a pilot plant. All 24 specimens are tested on a laboratory tensile tester in random order. The data from this experiment are shown in Table 13.1.

This is an example of a completely randomized single-factor experiment with four levels of the factor. The levels of the factor are sometimes called **treatments**, and each treatment has six observations or **replicates**. The role of **randomization** in this experiment is extremely important. By randomizing the order of the 24 runs, the effect of any nuisance variable that may influence the observed tensile strength is approximately balanced out. For example, suppose that there is a warm-up effect on the tensile testing machine; that is, the longer the machine is on, the greater the observed tensile strength. If all 24 runs are made in order of increasing hardwood concentration (that is, all six 5% concentration specimens are tested first, followed by all six 10% concentration specimens, etc.), any observed differences in tensile strength could also be due to the warm-up effect. The role of randomization to identify causality was discussed in Section 10.1.

It is important to graphically analyze the data from a designed experiment. Figure 13.1 presents box plots of tensile strength at the four hardwood concentration levels. This figure indicates that changing the hardwood concentration has an effect on tensile strength; specifically,

TABLE 13.1 **Tensile Strength of Paper (psi)**

Hardwood Concentration (%)	Observations						Totals	Averages
	1	2	3	4	5	6		
5	7	8	15	11	9	10	60	10.00
10	12	17	13	18	19	15	94	15.67
15	14	18	19	17	16	18	102	17.00
20	19	25	22	23	18	20	127	21.17
							383	15.96

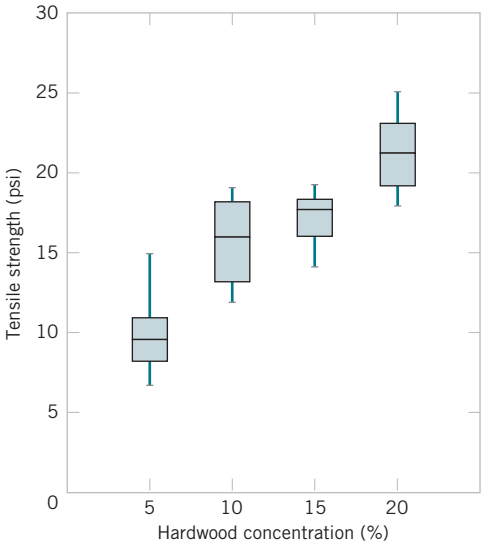


FIGURE 13.1
Box plots of hardwood concentration data.

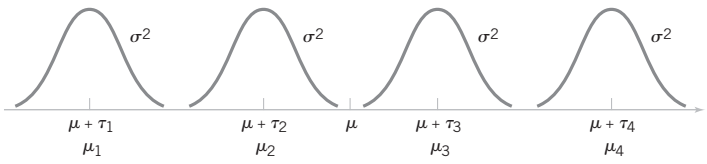


FIGURE 13.2
Display of the model for the completely randomized single-factor experiment.

higher hardwood concentrations produce higher observed tensile strength. Furthermore, the distribution of tensile strength at a particular hardwood level is reasonably symmetric, and the variability in tensile strength does not change dramatically as the hardwood concentration changes.

Graphical interpretation of the data is always useful. Box plots show the variability of the observations *within* a treatment (factor level) and the variability *between* treatments. We now discuss how the data from a single-factor randomized experiment can be analyzed statistically.

13.2.2 Analysis of Variance

Suppose that we have a different levels of a single factor that we wish to compare. Sometimes, each factor level is called a **treatment**, a very general term that can be traced to the early applications of experimental design methodology in the agricultural sciences. The *response* for each of the a treatments is a random variable. The observed data would appear as shown in Table 13.2. An entry in Table 13.2, say y_{ij}, represents the jth observation taken under treatment i. We initially consider the case that has an equal number of observations, n, on each of a treatments.

We may describe the observations in Table 13.2 by the *linear statistical model*

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases} \tag{13.1}$$

TABLE 13.2 Typical Data for a Single-Factor Experiment

Treatment	Observations				Totals	Averages
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots\vdots\vdots$	$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{an}	$y_{a\cdot}$	$\bar{y}_{a\cdot}$
					$y_{\cdot\cdot}$	$\bar{y}_{\cdot\cdot}$

where Y_{ij} is a random variable denoting the (ij)th observation, μ is a parameter common to all treatments called the **overall mean**, τ_i is a parameter associated with the ith treatment called the ith **treatment effect**, and ε_{ij} is a random error component. We assume that the errors ε_{ij} are normally and independently distributed with mean zero and variance σ^2. Notice that the model could have been written as

$$Y_{ij} = \mu_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases}$$

where $\mu_i = \mu + \tau_i$ is the mean of the ith treatment. In this form of the model, we see that each treatment defines a population that has mean μ_i consisting of the overall mean μ plus an effect τ_i that is due to that particular treatment. Therefore, each treatment can be thought of as a normal population with mean μ_i and variance σ^2. See Figure 13.2.

Furthermore, because we require that the observations are taken in random order and that the environment (often called the *experimental units*) in which the treatments are used is as uniform as possible, this experimental design is called a **completely randomized design (CRD)**.

We assume that the experimenter specifically chose the a treatments. In this situation, we wish to test hypotheses about the treatment means, and we may wish to estimate the treatment effects. This is called the **fixed-effects model**.

In this section, we develop the **analysis of variance (ANOVA)** for the fixed-effects model. The analysis of variance is not new to us; it was used previously in the presentation of regression analysis. However, in this section, we show how it can be used to test for equality of treatment effects. The treatment effects τ_i are usually defined as deviations from the overall mean μ, so that

$$\sum_{i=1}^{a} \tau_i = 0 \tag{13.2}$$

We are interested in testing the equality of the a treatment means $\mu_1, \mu_2, \ldots, \mu_a$. Using Equation 13.2, we find that this is equivalent to testing the hypotheses

$$H_0\colon \tau_1 = \tau_2 = \cdots = \tau_a = 0 \qquad H_1\colon \tau_i \neq 0 \text{ for at least one } i \tag{13.3}$$

Thus, if the null hypothesis is true, each observation consists of the overall mean μ plus a realization of the random error component ε_{ij}. This is equivalent to saying that all N observations are taken from a normal distribution with mean μ and variance σ^2. Therefore, if the null hypothesis is true, changing the levels of the factor has no effect on the mean response.

The analysis of variance partitions the total variability in the sample data into two component parts. Then, the test of the hypothesis in Equation 13.3 is based on a comparison of two independent estimates of the population variance. Let $y_{i\cdot}$ represent the total of the observations under the ith treatment and $\bar{y}_{i\cdot}$ represent the average of the observations under the ith treatment. Similarly, let $y_{\cdot\cdot}$ represent the grand total of all observations and $\bar{y}_{\cdot\cdot}$ represent the grand mean of all observations. Expressed mathematically,

$$\begin{aligned} y_{i\cdot} &= \sum_{j=1}^{n} y_{ij} & \bar{y}_{i\cdot} &= y_{i\cdot}/n & i &= 1, 2, \ldots, a \\ y_{\cdot\cdot} &= \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij} & \bar{y}_{\cdot\cdot} &= y_{\cdot\cdot}/N \end{aligned} \tag{13.4}$$

where $N = an$ is the total number of observations. Thus, the "dot" subscript notation implies summation over the subscript that it replaces.

The total variability in the data is described by the **total sum of squares**

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n} (y_{ij} - \bar{y}_{\cdot\cdot})^2$$

The partition of the total sum of squares is given in the following definition.

ANOVA Sum of Squares Identity: Single Factor Experiment

The **sum of squares identity** is

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 = n\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2 \tag{13.5}$$

or symbolically

$$SS_T = SS_{\text{Treatments}} + SS_E \tag{13.6}$$

and degrees of freedom can be partitioned as

$$an - 1 = a - 1 + a(n-1)$$

or

$$df_{\text{Total}} = df_{\text{Treatments}} + df_{\text{Error}}$$

The identity in Equation 13.5 shows that the total variability in the data can be partitioned into a sum of squares of differences between treatment means and the grand mean called the **treatment sum of squares**, and denoted $SS_{\text{Treatments}}$, and a sum of squares of differences of observations within a treatment from the treatment mean called the **error sum of squares**, and denoted SS_E. Differences between observed treatment means and the grand mean measure the differences between treatments, and differences of observations within a treatment from the treatment mean can be due only to random error.

There is also a partition of the number of degrees of freedom in Equation 13.6 that corresponds to the sum of squares identity. That is, there are $an = N$ observations; thus, SS_T has $an - 1$ degrees of freedom. There are a levels of the factor, so $SS_{\text{Treaments}}$ has $a - 1$ degrees of freedom. Finally, within any treatment, there are n replicates providing $n - 1$ degrees of freedom with which to estimate the experimental error. Because there are a treatments, we have $a(n - 1)$ degrees of freedom for error.

The ratio

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{a-1}$$

is called the **mean square for treatments** and the **mean square for error** is

$$MS_{\text{Error}} = \frac{SS_{\text{Error}}}{a(n-1)}$$

We can gain considerable insight into how the analysis of variance works by examining the expected values of $MS_{\text{Treatments}}$ and MS_E. This will lead us to an appropriate statistic for testing the hypothesis of no differences among treatment means (or all $\tau_i = 0$).

Expected Values of Mean Squares: Single Factor Experiment

The expected value of the treatment mean square is

$$E(MS_{\text{Treatments}}) = \sigma^2 + \frac{n\sum_{i=1}^{a}\tau_i^2}{a-1}$$

and the expected value of the error mean square is

$$E(MS_E) = \sigma^2$$

We can show that $MS_{\text{Treatments}}$ and MS_E are independent. Consequently, we can show that if the null hypothesis H_0 is true, the ratio

ANOVA *F*-Test

$$F_0 = \frac{SS_{\text{Treatments}}/(a-1)}{SS_E/[a(n-1)]} = \frac{MS_{\text{Treatments}}}{MS_E} \tag{13.7}$$

has an F-distribution with $a-1$ and $a(n-1)$ degrees of freedom. Furthermore, from the expected mean squares, we know that MS_E is an unbiased estimator of σ^2. Also, under the null hypothesis, $MS_{\text{Treatments}}$ is an unbiased estimator of σ^2. However, if the null hypothesis is false, the expected value of $MS_{\text{Treatments}}$ is greater than σ^2. Therefore, under the alternative hypothesis, the expected value of the numerator of the test statistic (Equation 13.7) is greater than the expected value of the denominator. Consequently, we reject H_0 if the statistic is large. This implies an upper-tailed, one-tailed critical region. Therefore, we reject H_0 if $f_0 > f_{\alpha,a-1,a(n-1)}$ where f_0 is the computed value of F_0 from Equation 13.7.

The computations for this test procedure are usually summarized in tabular form as shown in Table 13.3. This is called an **analysis of variance** (or **ANOVA**) **table**.

TABLE 13.3 Analysis of Variance for a Single-Factor Experiment, Fixed-Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	$SS_{\text{Treatments}}$	$a-1$	$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{a-1}$	$\frac{MS_{\text{Treatments}}}{MS_E}$
Error	SS_E	$a(n-1)$	$MS_{\text{Error}} = \frac{SS_{\text{Error}}}{a(n-1)}$	
Total	SS_T	$an-1$		

EXAMPLE 13.1 | Tensile Strength ANOVA

Consider the paper tensile strength experiment described in Section 13.2.1. This experiment is a completely randomized design. We can use the analysis of variance to test the hypothesis that different hardwood concentrations do not affect the mean tensile strength of the paper. The hypotheses are

$$H_0{:}\tau_1 = \tau_2 = \tau_3 = \tau_4 = 0 \qquad H_1{:}\tau_i \neq 0 \text{ for at least one } i$$

We use $\alpha = 0.01$. The sums of squares for the analysis of variance are computed as follows:

$$SS_T = 512.96$$
$$SS_{\text{Treatments}} = 382.79$$
$$SS_E = SS_T - SS_{\text{Treatments}} = 512.96 - 382.79 = 130.17$$

The ANOVA is summarized in Table 13.4. Because $f_{0.01,3,20} = 4.94$, we reject H_0 and conclude that hardwood concentration in the pulp significantly affects the mean strength of the paper. We can also find a P-value for this test statistic from computer software to be:

$$P(F_{3,20} > 19.60) \simeq 3.59 \times 10^{-6}$$

Computer software is used here to obtain the probability. Because the P-value is considerably smaller than $\alpha = 0.01$, we have strong evidence to conclude that H_0 is not true.

Practical Interpretation: There is strong evidence to conclude that hardwood concentration has an effect on tensile strength. However, the ANOVA does not tell us which levels of hardwood concentration result in different tensile strength means. We see how to answer this question in Section 13.2.3.

TABLE 13.4 ANOVA for the Tensile Strength Data

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-value
Hardwood concentration	382.79	3	127.60	19.60	3.59 E-6
Error	130.17	20	6.51		
Total	512.96	23			

It can also be useful to construct 95% confidence intervals (CIs) on each individual treatment mean. The mean of the ith treatment is defined as

$$\mu_i = \mu + \tau_i \quad i = 1, 2, \ldots, a$$

A point estimator of μ_i is $\hat{\mu}_i = \overline{Y}_{i\cdot}$. Now if we assume that the errors are normally distributed, each treatment average is normally distributed with mean μ_i and variance σ^2/n. Thus, if σ^2 were known, we could use the normal distribution to construct a CI. Using MS_E as an estimator of σ^2, we would base the CI on the t distribution because

$$T = \frac{\overline{Y}_{i\cdot} - \mu_i}{\sqrt{MS_E/n}}$$

has a t distribution with $a(n-1)$ degrees of freedom. This leads to the following definition of the confidence interval.

Confidence Interval on a Treatment Mean

A $100(1-\alpha)\%$ confidence interval on the mean of the ith treatment μ_i is

$$\overline{y}_{i\cdot} - t_{\alpha/2,a(n-1)}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \overline{y}_{i\cdot} + t_{\alpha/2,a(n-1)}\sqrt{\frac{MS_E}{n}} \tag{13.8}$$

For example, at 20% hardwood, the point estimate of the mean is $\overline{y}_{4\cdot} = 21.167$, $MS_E = 6.51$, and $t_{0.025,20} = 2.086$, so the 95% CI is

$$\overline{y}_{4\cdot} \pm t_{0.025,20}\sqrt{MS_E/n} = 21.167 \pm (2.086)\sqrt{6.51/6}$$

or

$$19.00 \text{ psi} \le \mu_4 \le 23.34 \text{ psi}$$

It can also be interesting to find confidence intervals on the difference in two treatment means, say, $\mu_i - \mu_j$. The point estimator of $\mu_i - \mu_j$ is $\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}$, and the variance of this estimator is

$$V(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}) = \frac{\sigma^2}{n} + \frac{\sigma^2}{n} = \frac{2\sigma^2}{n}$$

Now if we use MS_E to estimate σ^2,

$$T = \frac{\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot} - (\mu_i - \mu_j)}{\sqrt{2MS_E/n}}$$

has a t distribution with $a(n-1)$ degrees of freedom. Therefore, a CI on $\mu_i - \mu_j$ may be based on the T statistic.

Confidence Interval on a Difference in Treatment Means

A $100(1-\alpha)$ percent confidence interval on the difference in two treatment means $\mu_i - \mu_j$ is

$$\bar{y}_{i\cdot} - \bar{y}_{j\cdot} - t_{\alpha/2,a(n-1)}\sqrt{\frac{2MS_E}{n}} \le \mu_i - \mu_j \le \bar{y}_{i\cdot} - \bar{y}_{j\cdot} + t_{\alpha/2,a(n-1)}\sqrt{\frac{2MS_E}{n}} \tag{13.9}$$

A 95% CI on the difference in means $\mu_3 - \mu_2$ is computed from Equation 13.9 as follows:

$$\bar{y}_{3\cdot} - \bar{y}_{2\cdot} \pm t_{0.025,20}\sqrt{2MS_E/n} = 17.00 - 15.67 \pm (2.086)\sqrt{2(6.51)/6}$$

or

$$-1.74 \le \mu_3 - \mu_2 \le 4.40$$

Because the CI includes zero, we would conclude that there is no difference in mean tensile strength at these two particular hardwood levels.

Unbalanced Experiment In some single-factor experiments, the number of observations taken under each treatment may be different. We then say that the design is **unbalanced**. In this situation, slight modifications must be made in the sums of squares formulas. Let n_i observations be taken under treatment i ($i = 1, 2, \ldots, a$), and let the total number of observations $N = \sum_{i=1}^{a} n_i$. Computational formulas for SS_T and $SS_{\text{Treatments}}$ are as shown in the following definition.

Computing Formulas for ANOVA: Single Factor with Unequal Sample Sizes

The sums of squares computing formulas for the ANOVA with unequal sample sizes n_i in each treatment are

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{\cdot\cdot}^2}{N} \tag{13.10}$$

$$SS_{\text{Treatments}} = \sum_{i=1}^{a} \frac{y_{i\cdot}^2}{n_i} - \frac{y_{\cdot\cdot}^2}{N}$$

and

$$SS_E = SS_T - SS_{\text{Treatments}}$$

Choosing a balanced design has two important advantages. First, the ANOVA is relatively insensitive to small departures from the assumption of equality of variances if the sample sizes are equal. This is not the case for unequal sample sizes. Second, the power of the test is maximized if the samples are of equal size.

13.2.3 Multiple Comparisons Following the ANOVA

When the null hypothesis H_0:$\tau_1 = \tau_2 = \cdots = \tau_a = 0$ is rejected in the ANOVA, we know that some of the treatment or factor-level means are different. However, the ANOVA does not identify which means are different. Methods for investigating this issue are called **multiple comparisons methods**. Many of these procedures are available. Here we describe a very simple one, **Fisher's least significant difference (LSD) method** and a graphical method. Montgomery (2017) presents these and other methods and provides a comparative discussion.

The Fisher LSD method compares all pairs of means with the null hypotheses H_0:$\mu_i = \mu_j$ (for all $i \neq j$) using the t-statistic

$$t_0 = \frac{\bar{y}_{i\cdot} - \bar{y}_{j\cdot}}{\sqrt{\frac{2MS_E}{n}}}$$

Assuming a two-sided alternative hypothesis, the pair of means μ_i and μ_j would be declared significantly different if

$$|\bar{y}_{i\cdot} - \bar{y}_{j\cdot}| > \text{LSD}$$

where LSD, the least significant difference, is

Least Significant Difference for Multiple Comparisons

$$\text{LSD} = t_{\alpha/2,a(n-1)}\sqrt{\frac{2MS_E}{n}} \tag{13.11}$$

If the sample sizes are different in each treatment, the LSD is defined as

$$\text{LSD} = t_{\alpha/2,N-a}\sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

EXAMPLE 13.2

We apply the Fisher LSD method to the hardwood concentration experiment. There are $a = 4$ means, $n = 6$, $MS_E = 6.51$, and $t_{0.025,20} = 2.086$. The treatment means are

$$\bar{y}_{1\cdot} = 10.00 \text{ psi} \qquad \bar{y}_{2\cdot} = 15.67 \text{ psi}$$
$$\bar{y}_{3\cdot} = 17.00 \text{ psi} \qquad \bar{y}_{4\cdot} = 21.17 \text{ psi}$$

The value of LSD is $\text{LSD} = t_{0.025,20}\sqrt{2MS_E/n} = 2.086\sqrt{2(6.51)/6} = 3.07$. Therefore, any pair of treatment averages that differs by more than 3.07 implies that the corresponding pair of treatment means are different. The comparisons among the observed treatment averages are as follows:

4 vs. 1 = 21.17 − 10.00 = 11.17 > 3.07
4 vs. 2 = 21.17 − 15.67 = 5.50 > 3.07
4 vs. 3 = 21.17 − 17.00 = 4.17 > 3.07
3 vs. 1 = 17.00 − 10.00 = 7.00 > 3.07
3 vs. 2 = 17.00 − 15.67 = 1.33 < 3.07
2 vs. 1 = 15.67 − 10.00 = 5.67 > 3.07

Conclusions: From this analysis, we see that there are significant differences between all pairs of means except 2 and 3. This implies that 10% and 15% hardwood concentration produce approximately the same tensile strength and that all other concentration levels tested produce different tensile strengths. It is often helpful to draw a graph of the treatment means, such as in Figure 13.3 with the means that are *not* different underlined. This graph clearly reveals the results of the experiment and shows that 20% hardwood produces the maximum tensile strength.

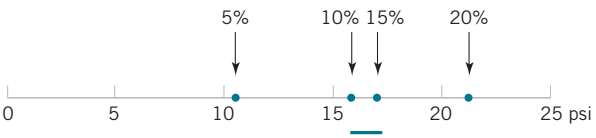


FIGURE 13.3
Results of Fisher's LSD method.

Graphical Comparison of Means It is easy to compare treatment means graphically following the analysis of variance. Suppose that the factor has a levels and that $\bar{y}_{1\cdot}, \bar{y}_{2\cdot}, \ldots, \bar{y}_{a\cdot}$ are the observed averages for these factor levels. Each treatment average has standard deviation $\sigma/\sqrt{n}$, where σ is the standard deviation of an individual observation. If all treatment means are equal,

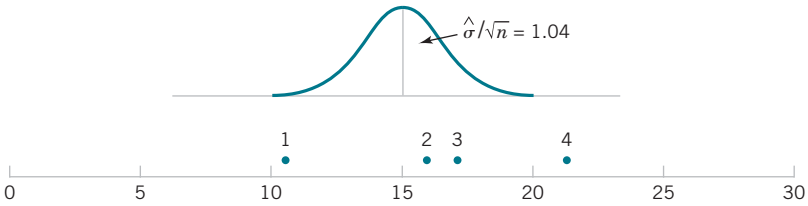


FIGURE 13.4

Tensile strength averages from the hardwood concentration experiment in relation to a normal distribution with standard deviation $\sqrt{MS_E/n} = \sqrt{6.51/6} = 1.04$.

the observed means $\bar{y}_{i\cdot}$ would behave as if they were a set of observations drawn at random from a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$.

Visualize this normal distribution capable of being slid along an axis below which the treatment means $\bar{y}_{1\cdot}, \bar{y}_{2\cdot}, \ldots, \bar{y}_{a\cdot}$ are plotted. If all treatment means are equal, there should be some position for this distribution that makes it obvious that the $\bar{y}_{i\cdot}$ values were drawn from the same distribution. If this is not the case, the $\bar{y}_{i\cdot}$ values that do not appear to have been drawn from this distribution are associated with treatments that produce different mean responses.

The only flaw in this logic is that σ is unknown. However, we can use $\sqrt{MS_E}$ from the analysis of variance to estimate σ. This implies that a t distribution should be used instead of a normal distribution in making the plot, but because the t looks so much like the normal, sketching a normal curve that is approximately $6\sqrt{MS_E/n}$ units wide will usually work very well.

Figure 13.4 shows this arrangement for the hardwood concentration experiment in Example 13.1. The standard deviation of this normal distribution is

$$\sqrt{MS_E/n} = \sqrt{6.51/6} = 1.04$$

If we visualize sliding this distribution along the horizontal axis, we note that there is no location for the distribution that would suggest that all four observations (the plotted means) are typical, randomly selected values from that distribution. This, of course, should be expected because the analysis of variance has indicated that the means differ, and the display in Figure 13.4 is just a graphical representation of the analysis of variance results. The figure does indicate that treatment 4 (20% hardwood) produces paper with higher mean tensile strength than do the other treatments, and treatment 1 (5% hardwood) results in lower mean tensile strength than do the other treatments. The means of treatments 2 and 3 (10% and 15% hardwood, respectively) do not differ.

This simple procedure is a rough but very effective multiple comparison technique. It works well in many situations.

13.2.4 Residual Analysis and Model Checking

The analysis of variance assumes that the observations are normally and independently distributed with the same variance for each treatment or factor level. These assumptions should be checked by examining the residuals. A **residual** is the difference between an observation y_{ij} and its estimated (or fitted) value from the statistical model being studied, denoted as $\hat{y}_{ij}$. For the completely randomized design $\hat{y}_{ij} = \bar{y}_{i\cdot}$ and each residual is

$$e_{ij} = y_{ij} - \bar{y}_{i\cdot}$$

This is the difference between an observation and the corresponding observed treatment mean. The residuals for the paper tensile strength experiment are shown in Table 13.5. Using $\bar{y}_{i\cdot}$ to calculate each residual essentially removes the effect of hardwood concentration from the data; consequently, the residuals contain information about unexplained variability.

The normality assumption can be checked by constructing a *normal probability plot* of the residuals. To check the assumption of equal variances at each factor level, plot the residuals against

TABLE 13.5 **Residuals for the Tensile Strength Experiment**

Hardwood Concentration (%)	Residuals					
5	−3.00	−2.00	5.00	1.00	−1.00	0.00
10	−3.67	1.33	−2.67	2.33	3.33	−0.67
15	−3.00	1.00	2.00	0.00	−1.00	1.00
20	−2.17	3.83	0.83	1.83	−3.17	−1.17

the factor levels and compare the spread in the residuals. It is also useful to plot the residuals against $\bar{y}_i$. (sometimes called the *fitted value*); the variability in the residuals should not depend in any way on the value of $\bar{y}_i$. Most statistical software packages can construct these plots on request. A pattern that appears in these plots usually suggests the need for a transformation, that is,

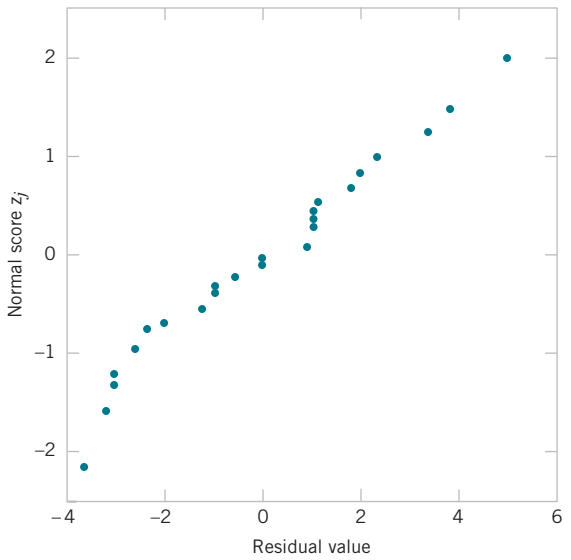


FIGURE 13.5(a)

Normal probability plot of residuals from the hardwood concentration experiment.

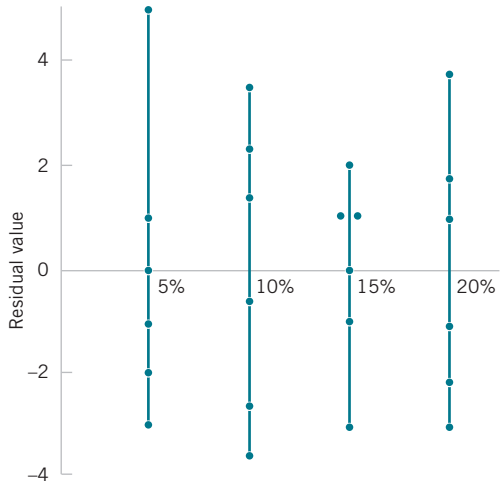


FIGURE 13.5(b)

Plot of residuals versus factor levels (hardwood concentration).

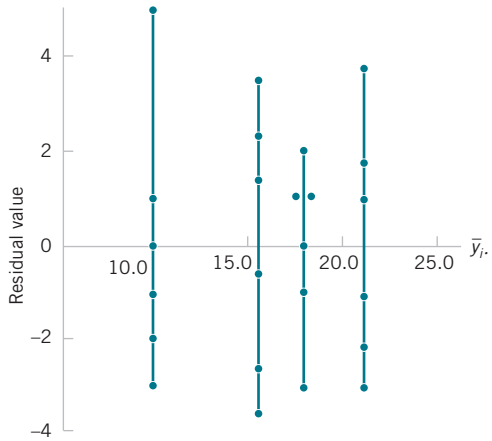


FIGURE 13.5(c)

Plot of residuals versus $\bar{y}_i$.

analyzing the data in a different metric. For example, if the variability in the residuals increases with $\bar{y}_i$., a transformation such as log y or $\sqrt{y}$ should be considered. In some problems, the dependency of residual scatter on the observed mean $\bar{y}_i$. is very important information. It may be desirable to select the factor level that results in maximum response; however, this level may also cause more variation in response from run to run.

The independence assumption can be checked by plotting the residuals against the time or run order in which the experiment was performed. A pattern in this plot, such as sequences of positive and negative residuals, may indicate that the observations are not independent. This suggests that time or run order is important or that variables that change over time are important and have not been included in the experimental design.

A normal probability plot of the residuals from the paper tensile strength experiment is shown in Figure 13.5(a). Figures 13.5(b) and 13.5(c) present the residuals plotted against the factor levels and the fitted value $\bar{y}_i$., respectively. These plots do not reveal any model inadequacy or unusual problem with the assumptions.

13.2.5 Determining Sample Size

In any experimental design problem, the choice of the sample size or number of replicates to use is important. **Operating characteristic (OC) curves** can be used to provide guidance here. Recall that an OC curve is a plot of the probability of a type II error (β) for various sample sizes against values of the parameters under test. The OC curves can be used to determine how many replicates are required to achieve adequate sensitivity.

The power of the ANOVA test is

$$\begin{aligned} 1-\beta &= P(\text{Reject } H_0 \mid H_0 \text{ is false}) \\ &= P(F_0 > f_{\alpha,a-1,a(n-1)} \mid H_0 \text{ is false}) \end{aligned} \tag{13.12}$$

To evaluate this probability statement, we need to know the distribution of the test statistic F_0 if the null hypothesis is false. Because ANOVA compares several means, the null hypothesis can be false in different ways. For example, possibly $\tau_1 > 0$, $\tau_2 = 0$, $\tau_3 < 0$, and so forth. It can be shown that the power for ANOVA in Equation 13.12 depends on the τ_i's only through the function

$$\Phi^2 = \frac{n\sum_{i=1}^{a}\tau_i^2}{a\sigma^2}$$

Therefore, an alternative hypotheses for the τ_i's can be used to calculate Φ^2 and this in turn can be used to calculate the power. Specifically, it can be shown that if H_0 is false, the statistic $F_0 = MS_{\text{Treatments}}/MS_E$ has a **noncentral *F*-distribution** with $a-1$ and $n(a-1)$ degrees of freedom and a noncentrality parameter that depends on Φ^2. Instead of tables for the noncentral F-distribution, OC curves are used to evaluate β defined in Equation 13.12. These curves plot β against Φ.

OC curves are available for $\alpha = 0.05$ and $\alpha = 0.01$ and for several values of the number of degrees of freedom for numerator (denoted v_1) and denominator (denoted v_2). Figure 13.6 gives representative OC curves, one for $a = 4$ ($v_1 = 3$) and one for $a = 5$ ($v_1 = 4$) treatments. Notice that for each value of a, there are curves for $\alpha = 0.05$ and $\alpha = 0.01$.

In using the curves, we must define the difference in means that we wish to detect in terms of $\sum_{i=1}^{a}\tau_i^2$. Also, the error variance σ^2 is usually unknown. In such cases, we must choose ratios of $\sum_{i=1}^{a}\tau_i^2/\sigma^2$ that we wish to detect. Alternatively, if an estimate of σ^2 is available, one may replace σ^2 with this estimate. For example, if we were interested in the sensitivity of an experiment that has already been performed, we might use MS_E as the estimate of σ^2.

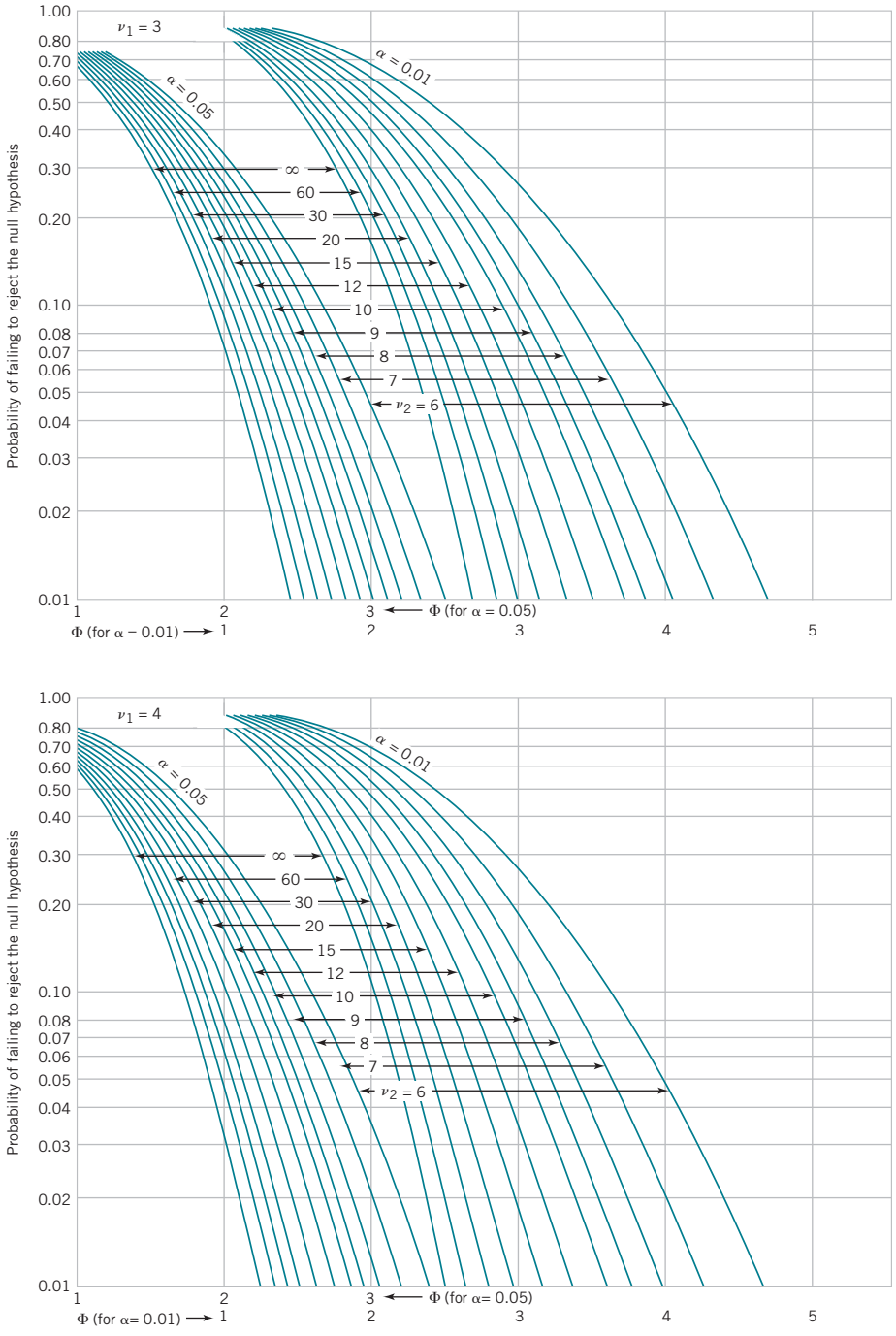


FIGURE 13.6

Operating characteristic curves for the fixed-effects model analysis of variance. Top curves for four treatments and bottom curves for five treatments.

EXAMPLE 13.3

Suppose that five means are being compared in a completely randomized experiment with $\alpha = 0.01$. The experimenter would like to know how many replicates to run if it is important to reject H_0 with probability at least 0.90 if $\sum_{i=1}^{5}\tau_i^2/\sigma^2 = 5.0$. The parameter Φ^2 is, in this case,

$$\Phi^2 = \frac{n\sum_{i=1}^{a}\tau_i^2}{a\sigma^2} = \frac{n}{5}(5) = n$$

and for the OC curve with $v_1 = a - 1 = 5 - 1 = 4$ and $v_2 = a(n-1) = 5(n-1)$ error degrees of freedom refer to the lower curve in Figure 13.6. As a first guess, try $n = 4$ replicates. This yields $\Phi^2 = 4$, $\Phi = 2$, and $v_2 = 5(3) = 15$ error degrees of freedom. Consequently, from Figure 13.6, we find that $\beta \simeq 0.38$. Therefore, the power of the test is approximately $1 - \beta = 1 - 0.38 = 0.62$, which is less than the required 0.90, so we conclude that $n = 4$ replicates is not sufficient. Proceeding in a similar manner, we can construct the following table:

n	Φ^2	Φ	$a(n-1)$	β	Power $= (1-\beta)$
4	4	2.00	15	0.38	0.62
5	5	2.24	20	0.18	0.82
6	6	2.45	25	0.06	0.94

Conclusions: At least $n = 6$ replicates must be run in order to obtain a test with the required power.

13.3 The Random-Effects Model

13.3.1 Fixed Versus Random Factors

In many situations, the factor of interest has a large number of possible levels. The analyst is interested in drawing conclusions about the entire population of factor levels. If the experimenter randomly selects a of these levels from the population of factor levels, we say that the factor is a **random factor**. Because the levels of the factor actually used in the experiment are chosen randomly, the conclusions reached are valid for the entire population of factor levels. We assume that the population of factor levels is either of infinite size or is large enough to be considered infinite. Notice that this is a very different situation than the one we encountered in the fixed-effects case in which the conclusions apply only for the factor levels used in the experiment.

13.3.2 ANOVA and Variance Components

The linear statistical model is

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,n \end{cases} \tag{13.13}$$

where the treatment effects τ_i and the errors ε_{ij} are independent random variables. Note that the model is identical in structure to the fixed-effects case, but the parameters have a different interpretation. If the variance of the treatment effects τ_i is σ_τ^2, by independence, the variance of the response is

$$V(Y_{ij}) = \sigma_\tau^2 + \sigma^2 \tag{13.14}$$

The variances σ_τ^2 and σ^2 are called **variance components**, and the model, Equation 13.14, is called the **components of variance model** or the **random-effects model**. To test hypotheses in this model, we assume that the errors ε_{ij} are normally and independently distributed with mean

zero and variance σ^2 and that the treatment effects τ_i are normally and independently distributed with mean zero and variance σ_τ^2.*

For the random-effects model, testing the hypothesis that the individual treatment effects are zero is meaningless. It is more appropriate to test hypotheses about σ_τ^2. Specifically,

$$H_0: \sigma_\tau^2 = 0 \qquad H_1: \sigma_\tau^2 > 0$$

If $\sigma_\tau^2 = 0$, all treatments are identical; but if $\sigma_\tau^2 > 0$, there is variability between treatments.

The ANOVA decomposition of total variability is still valid; that is,

$$SS_T = SS_{\text{Treatments}} + SS_E \tag{13.15}$$

However, the expected values of the mean squares for treatments and error are somewhat different than in the fixed-effects case.

Expected Values of Mean Squares: Random Effects

In the random-effects model for a single-factor, completely randomized experiment, the expected mean square for treatments is

$$E(MS_{\text{Treatments}}) = E\left(\frac{SS_{\text{Treatments}}}{a-1}\right) = \sigma^2 + n\sigma_\tau^2 \tag{13.16}$$

and the expected mean square for error is

$$E(MS_E) = E\left[\frac{SS_E}{a(n-1)}\right] = \sigma^2 \tag{13.17}$$

From examining the expected mean squares, it is clear that both MS_E and $MS_{\text{Treatments}}$ estimate σ^2 when $H_0: \sigma_\tau^2 = 0$ is true. Furthermore, MS_E and $MS_{\text{Treatments}}$ are independent. Consequently, the ratio

$$F_0 = \frac{MS_{\text{Treatments}}}{MS_E} \tag{13.18}$$

is an F random variable with $a - 1$ and $a(n - 1)$ degrees of freedom when H_0 is true. The null hypothesis would be rejected at the a-level of significance if the computed value of the test statistic $f_0 > f_{\alpha,a-1,a(n-1)}$.

The computational procedure and construction of the ANOVA table for the random-effects model are identical to the fixed-effects case. The conclusions, however, are quite different because they apply to the entire population of treatments.

Usually, we also want to estimate the variance components (σ^2 and σ_τ^2) in the model. The procedure that we use to estimate σ^2 and σ_τ^2 is called the **analysis of variance method** because it uses the information in the analysis of variance table. It does not require the normality assumption on the observations. The procedure consists of equating the expected mean squares to their observed values in the ANOVA table and solving for the variance components. When equating observed and expected mean squares in the one-way classification random-effects model, we obtain

$$MS_{\text{Treatments}} = \sigma^2 + n\sigma_\tau^2 \quad \text{and} \quad MS_E = \sigma^2$$

*The assumption that the $\{\tau_i\}$ are independent random variables implies that the usual assumption of $\sum_{i=1}^{n} \tau_i = 0$ from the fixed-effects model does not apply to the random-effects model.

Therefore, the estimators of the variance components are

ANOVA Variance Components Estimates

$$\hat{\sigma}^2 = MS_E \tag{13.19}$$

and

$$\hat{\sigma}_\tau^2 = \frac{MS_{\text{Treatments}} - MS_E}{n} \tag{13.20}$$

Sometimes the analysis of variance method produces a negative estimate of a variance component. Because variance components are by definition nonnegative, a negative estimate of a variance component is disturbing. One course of action is to accept the estimate and use it as evidence that the true value of the variance component is zero, assuming that sampling variation led to the negative estimate. Although this approach has intuitive appeal, it will disturb the statistical properties of other estimates. Another alternative is to reestimate the negative variance component with a method that always yields nonnegative estimates. Still another possibility is to consider the negative estimate as evidence that the assumed linear model is incorrect, requiring that a study of the model and its assumptions be made to find a more appropriate model.

EXAMPLE 13.4 | Textile Manufacturing

In *Design and Analysis of Experiments*, 8th edition (John Wiley, 2012), D. C. Montgomery describes a single-factor experiment involving the random-effects model in which a textile manufacturing company weaves a fabric on a large number of looms. The company is interested in loom-to-loom variability in tensile strength. To investigate this variability, a manufacturing engineer selects four looms at random and makes four strength determinations on fabric samples chosen at random from each loom. The data are shown in Table 13.6 and the ANOVA is summarized in Table 13.7.

TABLE 13.6 Strength Data

	Observations					
Loom	1	2	3	4	Total	Average
1	98	97	99	96	390	97.5
2	91	90	93	92	366	91.5
3	96	95	97	95	383	95.8
4	95	96	99	98	388	97.0
					1527	95.45

TABLE 13.7 Analysis of Variance for the Strength Data

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	*P*-value
Looms	89.19	3	29.73	15.68	1.88 E-4
Error	22.75	12	1.90		
Total	111.94	15			

From the analysis of variance, we conclude that the looms in the plant differ significantly in their ability to produce fabric of uniform strength. The variance components are estimated by $\hat{\sigma}^2 = 1.90$ and

$$\hat{\sigma}_\tau^2 = \frac{29.73 - 1.90}{4} = 6.96$$

Therefore, the variance of strength in the manufacturing process is estimated by

$$\hat{V}(Y_{ij}) = \hat{\sigma}_\tau^2 + \hat{\sigma}^2 = 6.96 + 1.90 = 8.86$$

Conclusion: Most of the variability in strength in the output product is attributable to differences between looms.

This example illustrates an important application of the analysis of variance—the isolation of different sources of variability in a manufacturing process. Problems of excessive variability in

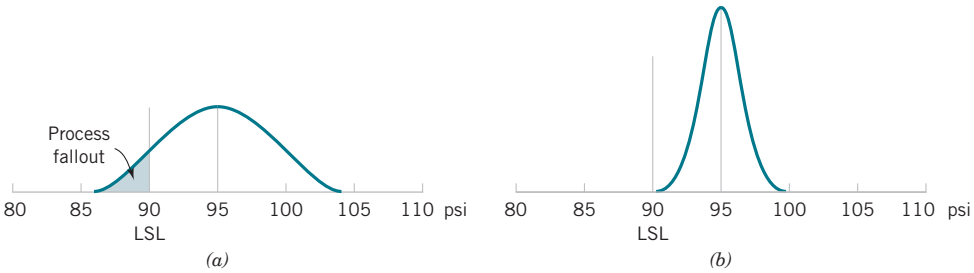


FIGURE 13.7

The distribution of fabric strength. (a) Current process, (b) Improved process.

critical functional parameters or properties frequently arise in quality-improvement programs. For example, in the previous fabric strength example, the process mean is estimated by $\bar{y} = 95.45$ psi, and the process standard deviation is estimated by $\hat{\sigma}_y = \sqrt{\hat{V}(Y_{ij})} = \sqrt{8.86} = 2.98$ psi. If strength is approximately normally distributed, the distribution of strength in the outgoing product would look like the normal distribution shown in Figure 13.7(a). If the lower specification limit (LSL) on strength is at 90 psi, a substantial proportion of the process output is **fallout**—that is, scrap or defective material that must be sold as second quality, and so on. This fallout is directly related to the excess variability resulting from differences between looms. Variability in loom performance could be caused by faulty setup, poor maintenance, inadequate supervision, poorly trained operators, and so forth. The engineer or manager responsible for quality improvement must identify and remove these sources of variability from the process. If this can be done, strength variability will be greatly reduced, perhaps as low as $\hat{\sigma}_Y = \sqrt{\hat{\sigma}^2} = \sqrt{1.90} = 1.38$ psi, as shown in Figure 13.7(b). In this improved process, reducing the variability in strength has greatly reduced the fallout, resulting in lower cost, higher quality, a more satisfied customer, and an enhanced competitive position for the company.

13.4 Randomized Complete Block Design

13.4.1 Design and Statistical Analysis

In many experimental design problems, it is necessary to design the experiment so that the variability arising from a nuisance factor can be controlled. In an earlier example with t-tests, two different methods were used to predict the shear strength of steel plate girders. Because each girder has different strength (potentially), and this variability in strength was not of direct interest, we designed the experiment by using the two test methods on each girder and then comparing the average difference in strength readings on each girder to zero using the paired t-test. The paired t-test is a procedure for comparing two treatment means when all experimental runs cannot be made under homogeneous conditions. Alternatively, we can view the paired t-test as a method for reducing the background noise in the experiment by blocking out a **nuisance factor** effect. The block is the nuisance factor, and, in this case, the nuisance factor is the steel girder specimens used in the experiment.

The randomized block design is an extension of the paired t-test to situations where the factor of interest has more than two levels; that is, more than two treatments must be compared. For example, suppose that three methods could be used to evaluate the strength readings on steel plate girders. We may think of these as three treatments, say t_1, t_2, and t_3. If we use four girders as the experimental units, a **randomized complete block design (RCBD)** would appear as shown in Figure 13.8. The design is called a RCBD because each block is large enough to hold all the treatments and because the actual assignment of each of the three treatments within each block is done randomly. Once the experiment has been conducted, the data are recorded in a table, such

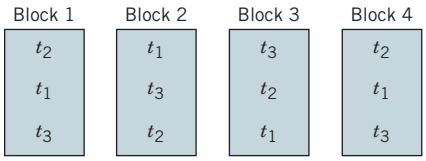


FIGURE 13.8
A randomized complete block design.

TABLE 13.8 A Randomized Complete Block Design

Treatments (Method)	Block (Girder) 1	2	3	4
1	y_{11}	y_{12}	y_{13}	y_{14}
2	y_{21}	y_{22}	y_{23}	y_{24}
3	y_{31}	y_{32}	y_{33}	y_{34}

as is shown in Table 13.8. The observations in this table, say, y_{ij}, represent the response obtained when method i is used on block j.

The general procedure for a RCBD consists of selecting b blocks and running a complete replicate of the experiment in each block. The data that result from running a RCBD for investigating a single factor with a levels and b blocks are shown in Table 13.9. There are a observations (one per factor level) in each block, and the order in which these observations are run is randomly assigned within the block.

We now describe the statistical analysis for the RCBD. Suppose that a single factor with a levels is of interest and that the experiment is run in b blocks. The observations may be represented by the *linear statistical model*

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \end{cases} \tag{13.21}$$

where μ is an overall mean, τ_i is the effect of the ith treatment, β_j is the effect of the jth block, and ϵ_{ij} is the random error term, which is assumed to be normally and independently distributed with mean zero and variance σ^2. Furthermore, the treatment and block effects are defined as deviations from the overall mean, so $\sum_{i=1}^{a} \tau_i = 0$ and $\sum_{j=1}^{b} \beta_j = 0$. This was the same type of definition used for completely randomized experiments in Section 13.2. We also assume that treatments and blocks do not interact. That is, the effect of treatment i is the same regardless of which block (or blocks) in which it is tested. We are interested in testing the equality of the treatment effects. That is,

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0 \qquad H_1: \tau_i \neq 0 \text{ at least one } i$$

The analysis of variance can be extended to the RCBD. The procedure uses a sum of squares identity that partitions the total sum of squares into three components.

TABLE 13.9 A Randomized Complete Block Design with a Treatments and b Blocks

Treatments	Blocks 1	2	…	b	Totals	Averages
1	y_{11}	y_{12}	…	y_{1b}	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	…	y_{2b}	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
⋮	⋮	⋮		⋮	⋮	⋮
a	y_{a1}	y_{a2}	…	y_{ab}	$y_{a\cdot}$	$\bar{y}_{a\cdot}$
Totals	$y_{\cdot 1}$	$y_{\cdot 2}$	…	$y_{\cdot b}$	$y_{\cdot\cdot}$	
Averages	$\bar{y}_{\cdot 1}$	$\bar{y}_{\cdot 2}$	…	$\bar{y}_{\cdot b}$		$\bar{y}_{\cdot\cdot}$

ANOVA Sums of Squares Identity: Randomized Complete Block Experiment

The sum of squares identity for the randomized complete block design is

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}..)^2 = b\sum_{i=1}^{a}(\bar{y}_i.-\bar{y}..)^2 + a\sum_{j=1}^{b}(\bar{y}_{\cdot j}-\bar{y}..)^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{\cdot j}-\bar{y}_i.+\bar{y}..)^2 \tag{13.22}$$

or symbolically

$$SS_T = SS_{\text{Treatments}} + SS_{\text{Blocks}} + SS_E$$

and degrees of freedom can be partitioned as

$$ab-1 = (a-1)+(b-1)+(a-1)(b-1)$$

or

$$df_{\text{Total}} = df_{\text{Treatments}} + df_{\text{Blocks}} + df_{\text{Error}}$$

For the randomized block design, the relevant mean squares are

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{a-1}$$

$$MS_{\text{Blocks}} = \frac{SS_{\text{Blocks}}}{b-1}$$

$$MS_E = \frac{SS_E}{(a-1)(b-1)}$$

The expected values of these mean squares can be shown to be as follows:

Expected Mean Squares: Randomized Complete Block Experiment

$$E(MS_{\text{Treatments}}) = \sigma^2 + \frac{b\sum_{i=1}^{a}\tau_i^2}{a-1}$$

$$E(MS_{\text{Blocks}}) = \sigma^2 + \frac{a\sum_{j=1}^{b}\beta_j^2}{b-1}$$

$$E(MS_E) = \sigma^2$$

Therefore, if the null hypothesis H_0 is true so that all treatment effects $\tau_i = 0$, $MS_{\text{Treatments}}$ is an unbiased estimator of σ^2, and if H_0 is false, $MS_{\text{Treatments}}$ overestimates σ^2. The mean square for error is always an unbiased estimate of σ^2. To test the null hypothesis that the treatment effects are all zero, we use the ratio

$$F_0 = \frac{MS_{\text{Treatments}}}{MS_E} \tag{13.23}$$

TABLE 13.10 ANOVA for a Randomized Complete Block Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	$SS_{\text{Treatments}}$	$a-1$	$MS_{\text{Treatments}} = \dfrac{SS_{\text{Treatments}}}{a-1}$	$\dfrac{MS_{\text{Treatments}}}{MS_E}$
Blocks	SS_{Blocks}	$b-1$	$MS_{\text{Blocks}} = \dfrac{SS_{\text{Blocks}}}{b-1}$	
Error	SS_E (by subtraction)	$(a-1)(b-1)$	$MS_{\text{Error}} = \dfrac{SS_{\text{Error}}}{(a-1)(b-1)}$	
Total	SS_T	$ab-1$		

which has an F-distribution with $a-1$ and $(a-1)(b-1)$ degrees of freedom if the null hypothesis is true. We would reject the null hypothesis at the α-level of significance if the computed value of the test statistic in Equation 13.23 is $f_0 > f_{\alpha,a-1,(a-1)(b-1)}$.

In practice, we compute SS_T, $SS_{\text{Treatments}}$ and SS_{Blocks} and then obtain the error sum of squares SS_E by subtraction. The appropriate computing formulas are as follows.

Computing Formulas for ANOVA: Randomized Block Experiment

The computing formulas for the sums of squares in the analysis of variance for a RCBD are

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{ab} \tag{13.24}$$

$$SS_{\text{Treatments}} = \frac{1}{b}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{ab} \tag{13.25}$$

$$SS_{\text{Blocks}} = \frac{1}{a}\sum_{j=1}^{b} y_{.j}^2 - \frac{y_{..}^2}{ab} \tag{13.26}$$

and

$$SS_E = SS_T - SS_{\text{Treatments}} - SS_{\text{Blocks}}$$

The computations are usually arranged in an ANOVA table, such as is shown in Table 13.10. Generally, computer software is used to perform the analysis of variance for a RCBD.

EXAMPLE 13.5 | Fabric Strength

An experiment was performed to determine the effect of four different chemicals on the strength of a fabric. These chemicals are used as part of the permanent press finishing process. Five fabric samples were selected, and a RCBD was run by testing each chemical type once in random order on each fabric sample. The data are shown in Table 13.11. We test for differences in means using an ANOVA with $\alpha = 0.01$.

The sums of squares for the analysis of variance are computed as follows:

$$SS_T = 25.69 \quad SS_{\text{Treatments}} = 18.04 \quad SS_{\text{Blocks}} = 6.69$$

$$\begin{aligned} SS_E &= SS_T - SS_{\text{Blocks}} - SS_{\text{Treatments}} \\ &= 25.69 - 6.69 - 18.04 = 0.96 \end{aligned}$$

The ANOVA is summarized in Table 13.12. Because $f_0 = 75.13 > f_{0.01,3,12} = 5.95$ (the P-value is 4.79×10^{-8} from computer software), we conclude that there is a significant difference in the chemical types so far as their effect on strength is concerned.

TABLE 13.11 Fabric Strength Data—Randomized Complete Block Design

Chemical Type	Fabric Sample 1	2	3	4	5	Treatment Totals $y_{i\cdot}$	Treatment Averages $\bar{y}_{i\cdot}$
1	1.3	1.6	0.5	1.2	1.1	5.7	1.14
2	2.2	2.4	0.4	2.0	1.8	8.8	1.76
3	1.8	1.7	0.6	1.5	1.3	6.9	1.38
4	3.9	4.4	2.0	4.1	3.4	17.8	3.56
Block totals $y_{\cdot j}$	9.2	10.1	3.5	8.8	7.6	39.2($y_{\cdot\cdot}$)	
Block averages $\bar{y}_{\cdot j}$	2.30	2.53	0.88	2.20	1.90		1.96($\bar{y}_{\cdot\cdot}$)

TABLE 13.12 Analysis of Variance for the Randomized Complete Block Experiment

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-value
Chemical types (treatments)	18.04	3	6.01	75.13	4.79 E-8
Fabric samples (blocks)	6.69	4	1.67		
Error	0.96	12	0.08		
Total	25.69	19			

When Is Blocking Necessary? Suppose that an experiment is conducted as a RCBD and blocking was not really necessary. There are ab observations and $(a-1)(b-1)$ degrees of freedom for error. If the experiment had been run as a completely randomized single-factor design with b replicates, we would have had $a(b-1)$ degrees of freedom for error. Therefore, blocking has cost $a(b-1)-(a-1)(b-1)=b-1$ degrees of freedom for error. This loss in degrees of freedom increases the critical value from the F-distribution in Equation 13.23. Consequently, to detect a treatment effect, greater differences in treatment means are needed. However, because the loss in error degrees of freedom is usually small, if there is a reasonable chance that block effects may be important, the experimenter should use the RCBD.

For example, consider the experiment described in Example 13.5 as a single-factor experiment with no blocking. We would then have 16 degrees of freedom for error. In the RCBD, there are 12 degrees of freedom for error. Therefore, blocking has cost only 4 degrees of freedom, which is a very small loss considering the possible gain in information that would be achieved if block effects are really important. The block effect in Example 13.5 is large, and if we had not blocked, SS_{Blocks} would have been included in the error sum of squares for the completely randomized analysis. This would have resulted in a much larger MS_E, making it more difficult to detect treatment differences. As a general rule, when in doubt as to the importance of block effects, the experimenter should assume that the block effect does exist. If the experimenter is wrong, the slight loss in the degrees of freedom for error will have a negligible effect unless the number of degrees of freedom is very small.

13.4.2 Multiple Comparisons

When the ANOVA indicates that a difference for the treatment means exists, we may need to perform some follow-up tests to isolate the specific differences. Any multiple comparison method could be used for this purpose. We illustrate Fisher's LSD method. The four chemical type averages from Example 13.5 are:

$$\bar{y}_{1\cdot}=1.14 \qquad \bar{y}_{2\cdot}=1.76 \qquad \bar{y}_{3\cdot}=1.38 \qquad \bar{y}_{4\cdot}=3.56$$

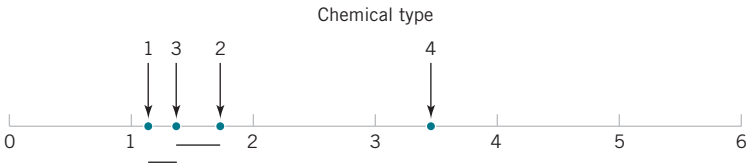


FIGURE 13.9
Results of Fisher's LSD method.

Each treatment average uses $b = 5$ observations (one from each block). We use $\alpha = 0.05$, so $t_{0.025,12} = 2.179$. Therefore, the value of the LSD is

$$\text{LSD} = t_{0.025,12}\sqrt{\frac{2MS_E}{b}} = 2.179\sqrt{\frac{2(0.08)}{5}} = 0.39$$

Any pair of treatment averages that differ by 0.39 or more indicates that this pair of treatment means is significantly different. The comparisons follow:

$$4 \text{ vs. } 1 = \bar{y}_4. - \bar{y}_1. = 3.56 - 1.14 = 2.42 > 0.39$$
$$4 \text{ vs. } 3 = \bar{y}_4. - \bar{y}_3. = 3.56 - 1.38 = 2.18 > 0.39$$
$$4 \text{ vs. } 2 = \bar{y}_4. - \bar{y}_2. = 3.56 - 1.76 = 1.80 > 0.39$$
$$2 \text{ vs. } 1 = \bar{y}_2. - \bar{y}_1. = 1.76 - 1.14 = 0.62 > 0.39$$
$$2 \text{ vs. } 3 = \bar{y}_2. - \bar{y}_3. = 1.76 - 1.38 = 0.38 < 0.39$$
$$3 \text{ vs. } 1 = \bar{y}_3. - \bar{y}_1. = 1.38 - 1.14 = 0.24 < 0.39$$

Figure 13.9 presents the results graphically. The underlined pairs of means are *not* different. The LSD procedure indicates that chemical type 4 results in significantly different strengths than the other three types do. Chemical types 2 and 3 do not differ, and types 1 and 3 do not differ. There may be a small difference in strength for types 1 and 2.

13.4.3 Residual Analysis and Model Checking

In any designed experiment, it is always important to examine the residuals and to check for violation of basic assumptions that could invalidate the results. As usual, the residuals for the RCBD are simply the differences in the observed and estimated (or fitted) values from the statistical model, $e_{ij} = y_{ij} - \hat{y}_{ij}$, and the fitted values are

$$\hat{y}_{ij} = \bar{y}_i. + \bar{y}._j - \bar{y}..$$

The fitted value represents the estimate of the mean response when the ith treatment is run in the jth block. The residuals from the chemical type experiment are shown in Table 13.13.

TABLE 13.13 Residuals from the Randomized Complete Block Design

	Fabric Sample				
Chemical Type	1	2	3	4	5
1	−0.18	−0.10	0.44	−0.18	0.02
2	0.10	0.08	−0.28	0.00	0.10
3	0.08	−0.24	0.30	−0.12	−0.02
4	0.00	0.28	−0.48	0.30	−0.10

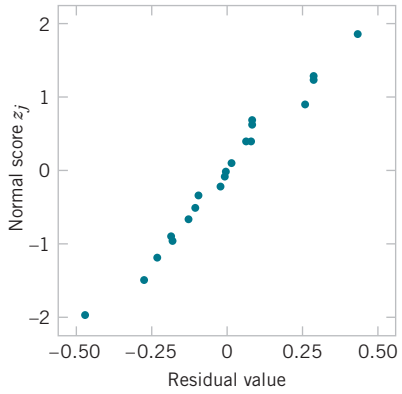


FIGURE 13.10(a)

Normal probability plot of residuals from the randomized complete block design.

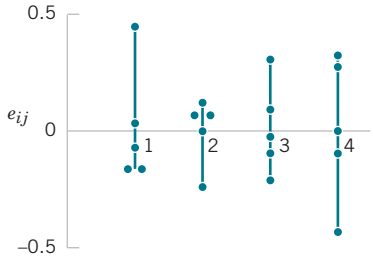


FIGURE 13.10(b)

Residuals by treatment from the randomized complete block design.

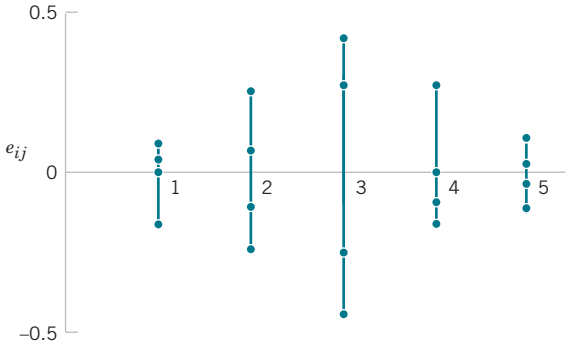


FIGURE 13.10(c)

Residuals by block from the randomized complete block design.

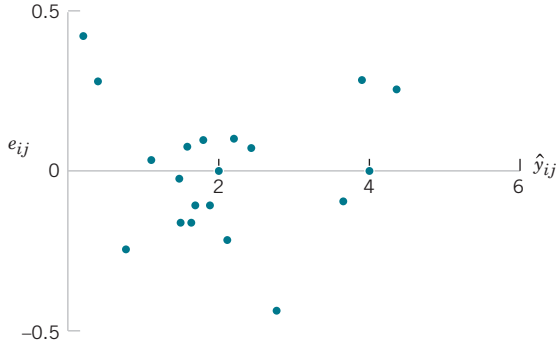


FIGURE 13.10(d)

Residuals versus $\hat{y}_{ij}$ from the randomized complete block design.

Figures 13.10(a)–(d) present the important residual plots for the experiment. These residual plots are usually constructed by computer software. When treated with the four chemicals, there is some indication that fabric sample (block) 3 has greater variability in strength than the other samples. Chemical type 4, which provides the greatest strength, also has somewhat more variability in strength. Follow-up experiments may be necessary to confirm these findings if they are potentially important.

Important Terms and Concepts

Analysis of variance (ANOVA)
Blocking
Completely randomized design (CRD)
Components of variance model
Error mean square
Fisher's least significant difference (LSD) method
Fixed-effects model
Graphical comparison of means
Least significant difference
Levels of a factor
Multiple comparisons methods
Nuisance factor
Operating characteristic (OC) curves
Random factor
Random-effects model
Randomization
Randomized complete block design (RCBD)
Residual analysis and model checking
Treatment
Treatment mean square
Variance components

CHAPTER **14**

Design of Experiments with Several Factors

© RedHelga / iStockphoto

LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Design and conduct engineering experiments involving several factors using the factorial design approach
2. Know how to analyze and interpret main effects and interactions
3. Understand how to use the ANOVA to analyze the data from these experiments
4. Assess model adequacy with residual plots
5. Know how to use the two-level series of factorial designs
6. Understand how to run two-level factorial design in blocks
7. Design and conduct two-level fractional factorial designs
8. Use center points to test for curvature in two-level factorial designs
9. Use response surface methodology for process optimization experiments

CHAPTER OUTLINE

14.1 Introduction

14.2 Factorial Experiments

14.3 Two-Factor Factorial Experiments

- **14.3.1** Statistical Analysis
- **14.3.2** Model Adequacy Checking
- **14.3.3** One Observation per Cell

14.4 General Factorial Experiments

14.5 2^k Factorial Designs

- **14.5.1** 2^2 Design
- **14.5.2** 2^k Design for $k \geq 3$ Factors

14.6 Single Replicate of the 2^k Design

14.7 Addition of Center Points to a 2^k Design

14.8 Blocking and Confounding in the 2^k Design

14.9 One-Half Fraction of the 2^k Design

14.10 Smaller Fractions: The 2^{k-p} Fractional Factorial

14.11 Response Surface Methods and Designs

Carotenoids are fat-soluble pigments that occur naturally in fruits and vegetables and are recommended for healthy diets. A well-known carotenoid is beta-carotene. Astaxanthin is another carotenoid that is a strong antioxidant and is commercially produced. An exercise later in this chapter describes an experiment in *Biotechnology Progress* to promote astaxanthin production. Seven variables were considered important to production: photon flux density and concentrations of nitrogen, phosphorous, magnesium, acetate, ferrous, and NaCl. It was important to study the effects of these factors as well as the effects of combinations on the production. Even with only a high and low setting for each variable, an experiment that uses all possible combinations requires $2^7 = 128$ tests. Such a large experiment has a number of disadvantages, and a question is whether a fraction of the full set of tests can be selected to provide the most important information in many fewer runs. The example used a surprisingly small set of 16 runs (16/128 = 1/8 fraction). The design and analysis of experiments of this type is the focus of this chapter. Such experiments are widely used throughout modern engineering development and scientific studies.

14.1 Introduction

An **experiment** is just a test or series of tests. Experiments are performed in all engineering and scientific disciplines and are important parts of the way we learn about how systems and processes work. The validity of the conclusions that are drawn from an experiment depends to a large extent on how the experiment was conducted. Therefore, the **design** of the experiment plays a major role in the eventual solution to the problem that initially motivated the experiment.

In this chapter, we focus on experiments that include two or more factors that the experimenter thinks may be important. A **factorial experiment** is a powerful technique for this type of problem. Generally, in a factorial experimental design, experimental trials (or runs) are performed at all combinations of factor levels. For example, if a chemical engineer is interested in investigating the effects of reaction time and reaction temperature on the yield of a process, and if two levels of time (1.0 and 1.5 hours) and two levels of temperature (125 and 150°F) are considered important, a factorial experiment would consist of making experimental runs at each of the four possible combinations of these levels of reaction time and reaction temperature.

Experimental design is an extremely important tool for engineers and scientists who are interested in improving the performance of a manufacturing process. It also has extensive application in the development of new processes and in new product design. We now give some examples.

Process Characterization Experiment In an article in *IEEE Transactions* ["Electronics Packaging Manufacturing" (2001, Vol. 24(4), pp. 249–254)], the authors discussed the change to lead-free solder in surface mount technology (SMT). SMT is a process to assemble electronic components to a printed circuit board. Solder paste is printed through a stencil onto the printed circuit board. The stencil-printing machine has squeegees; the paste rolls in front of the squeegee and fills the apertures in the stencil. The squeegee shears off the paste in the apertures as it moves over the stencil. Once the print stroke is completed, the board is separated mechanically from the stencil. Electronic components are placed on the deposits, and the board is heated so that the paste reflows to form the solder joints.

The process has several (perhaps many) variables, and all of them are not equally important. The initial list of candidate variables to be included in an experiment is constructed by combining the knowledge and information about the process from all team members. For example, engineers would conduct a *brainstorming* session and invite manufacturing personnel with SMT experience to participate. SMT has several variables that can be controlled. These include (1) squeegee speed,

(2) squeegee pressure, (3) squeegee angle, (4) metal or polyurethane squeegee, (5) squeegee vibration, (6) delay time before the squeegee lifts from the stencil, (7) stencil separation speed, (8) print gap, (9) solder paste alloy, (10) paste pretreatment (11) paste particle size, (12) flux type, (13) reflow temperature, (14) reflow time, and so forth.

In addition to these controllable factors, several other factors cannot be easily controlled during routine manufacturing, including (1) thickness of the printed circuit board, (2) types of components used on the board and aperture width and length, (3) layout of the components on the board, (4) paste density variation, (5) environmental factors, (6) squeegee wear, (7) cleanliness, and so forth. Sometimes we call the uncontrollable factors *noise* factors. A schematic representation of the process is shown in Figure 14.1. In this situation, the engineer wants to **characterize** the SMT process, that is, to determine the factors (both controllable and uncontrollable) that affect the occurrence of defects on the printed circuit boards. To determine these factors, an experiment can be designed to estimate the magnitude and direction of the factor effects. Sometimes we call such an experiment a **screening experiment**. The information from this characterization study, or screening experiment, can help determine the critical process variables as well as the direction of adjustment for these factors to reduce the number of defects.

Optimization Experiment In a characterization experiment, we are interested in determining which factors affect the response. A logical next step is to determine the region in the important factors that leads to an optimum response. For example, if the response is cost, we look for a region of minimum cost. This leads to an **optimization experiment**.

As an illustration, suppose that the yield of a chemical process is influenced by the operating temperature and the reaction time. We are currently operating the process at 155°F and 1.7 hours of reaction time, and the current process yield is around 75%. See Figure 14.2 for a view of the time–temperature space. In this graph, we have connected points of constant yield with lines. These lines are yield *contours*, and we have shown the contours at 60, 70, 80, 90, and 95% yield. To locate the optimum, we might begin with a factorial experiment such as we describe here, with the two factors, time and temperature, run at two levels each at 10°F and 0.5 hours above and below the current operating conditions. This two-factor factorial design is shown in Figure 14.2. The average responses observed at the four points in the experiment (145°F, 1.2 hours; 145°F, 2.2 hours; 165°F, 1.2 hours; and 165°F, 2.2 hours) indicate that we should move in the general direction of increased temperature and lower reaction time to increase yield. A few additional runs could be performed in this direction to locate the region of maximum yield.

Product Design Example We can also use experimental design in developing new products. For example, suppose that a group of engineers are designing a door hinge for an automobile. The product characteristic is the check effort, or the holding ability, of the latch that

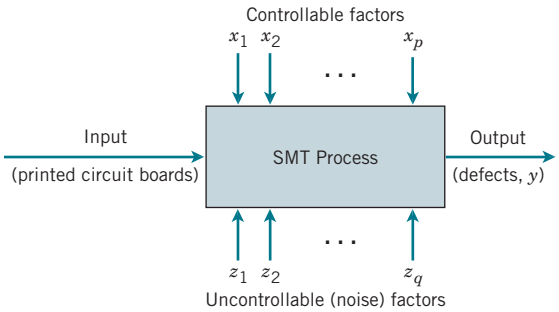


FIGURE 14.1
The flow solder experiment.

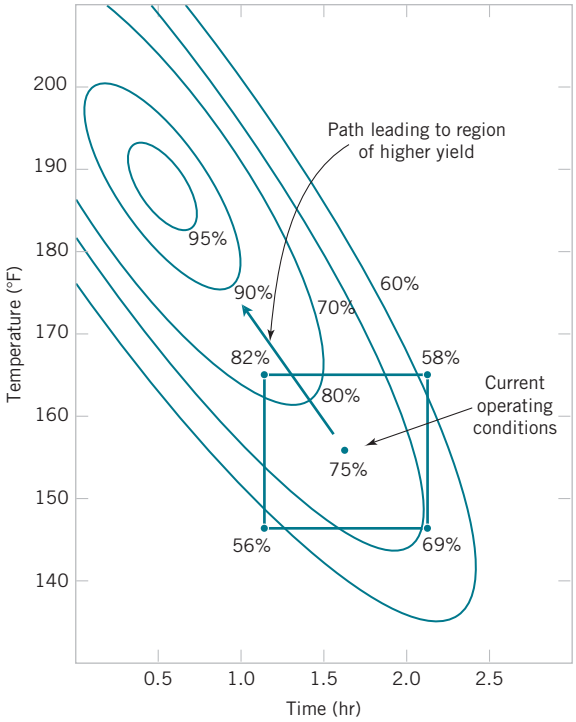


FIGURE 14.2

Contour plot of yield as a function of reaction time and reaction temperature, illustrating an optimization experiment.

prevents the door from swinging closed when the vehicle is parked on a hill. The check mechanism consists of a leaf spring and a roller. When the door is opened, the roller travels through an arc causing the leaf spring to be compressed. To close the door, the spring must be forced aside, creating the check effort. The engineering team thinks that check effort is a function of the following factors: (1) roller travel distance, (2) spring height from pivot to base, (3) horizontal distance from pivot to spring, (4) free height of the reinforcement spring, and (5) free height of the main spring.

The engineers can build a prototype hinge mechanism in which all these factors can be varied over certain ranges. Once appropriate levels for these five factors have been identified, the engineers can design an experiment consisting of various combinations of the factor levels and can test the prototype at these combinations. This produces information concerning which factors are most influential on the latch check effort, and through analysis of this information, the latch design can be improved.

14.2 Factorial Experiments

When several factors are of interest in an experiment, a factorial experiment should be used. As noted previously, in these experiments factors are varied together.

Factorial Experiment

By **factorial experiment**, we mean that in each complete trial or replicate of the experiment, all possible combinations of the levels of the factors are investigated.

Thus, if there are two factors A and B with a levels of factor A and b levels of factor B, each replicate contains all ab treatment combinations.

The effect of a factor is defined as the change in response produced by a change in the level of the factor. It is called a **main effect** because it refers to the primary factors in the study. For example, consider the data in Table 14.1. This is a factorial experiment with two factors, A and B, each at two levels (A_{low}, A_{high} and B_{low}, B_{high}). The main effect of factor A is the difference between the average response at the high level of A and the average response at the low level of A, or

$$A = \frac{30 + 40}{2} - \frac{10 + 20}{2} = 20$$

That is, changing factor A from the low level to the high level causes an average response increase of 20 units. Similarly, the main effect of B is

$$B = \frac{20 + 40}{2} - \frac{10 + 30}{2} = 10$$

In some experiments, the difference in response between the levels of one factor is not the same at all levels of the other factors. When this occurs, there is an **interaction** between the factors. For example, consider the data in Table 14.2. At the low level of factor B, the A effect is

$$A = 30 - 10 = 20$$

and at the high level of factor B, the A effect is

$$A = 0 - 20 = -20$$

Because the effect of A depends on the level chosen for factor B, there is interaction between A and B.

When an interaction is large, the corresponding main effects have very little practical meaning. For example, by using the data in Table 14.2, we find the main effect of A as

$$A = \frac{30 + 0}{2} - \frac{10 + 20}{2} = 0$$

and we would be tempted to conclude that there is no factor A effect. However, when we examined the effects of A at *different levels of factor* B, we saw that this was not the case. The effect of factor A depends on the levels of factor B. Thus, knowledge of the AB interaction is more useful than knowledge of the main effect. A significant interaction can mask the significance of main effects. Consequently, when interaction is present, the main effects of the factors involved in the interaction may not have much meaning.

It is easy to estimate the interaction effect in factorial experiments such as those illustrated in Tables 14.1 and 14.2. In this type of experiment, when both factors have two levels, the AB interaction effect is the difference in the diagonal averages. This represents one-half the difference

TABLE 14.1 **A Factorial Experiment with Two Factors**

	Factor B	
Factor A	B_{low}	B_{high}
A_{low}	10	20
A_{high}	30	40

TABLE 14.2 **A Factorial Experiment with Interaction**

	Factor B	
Factor A	B_{low}	B_{high}
A_{low}	10	20
A_{high}	30	0

between the A effects at the two levels of B. For example, in Table 14.1, we find the AB interaction effect to be

$$AB = \frac{20 + 30}{2} - \frac{10 + 40}{2} = 0$$

Thus, there is no interaction between A and B. In Table 14.2, the AB interaction effect is

$$AB = \frac{20 + 30}{2} - \frac{10 + 0}{2} = 20$$

As we noted before, the interaction effect in these data is very large.

The concept of interaction can be illustrated graphically in several ways. See Figure 14.3, which plots the data in Table 14.1 against the levels of A for both levels of B. Note that the B_{low} and B_{high} lines are approximately parallel, indicating that factors A and B do not interact significantly. Figure 14.4 presents a similar plot for the data in Table 14.2. In this graph, the B_{low} and B_{high} lines are not parallel, indicating the interaction between factors A and B. Such graphical displays are called *two-factor interaction plots*. They are often useful in presenting the results of experiments, and many computer software programs used for analyzing data from designed experiments construct these graphs automatically.

Figures 14.5 and 14.6 present two other graphical illustrations of the data from Tables 14.1 and 14.2. In Figure 14.3, we have shown a *three-dimensional surface plot* of the data from Table 14.1. These data contain no interaction, and the surface plot is a plane lying above the A-B space. The slope of the plane in the A and B directions is proportional to the main effects of factors A and B, respectively. Figure 14.6 is a surface plot of the data from Table 14.2. Notice that the effect of the interaction in these data is to "twist" the plane so that there is curvature in the response function. **Factorial experiments are the only way to discover interactions between variables**.

An alternative to the factorial design that is (unfortunately) used in practice is to change the factors *one at a time* rather than to vary them simultaneously. To illustrate this one-factor-at-a-time procedure, suppose that an engineer is interested in finding the values of temperature and pressure that maximize yield in a chemical process. Suppose that we fix temperature at 155°F (the current operating level) and perform five runs at different levels of time, say, 0.5, 1.0, 1.5, 2.0, and 2.5 hours. The results of this series of runs are shown in Figure 14.7. This figure indicates that maximum yield is achieved at about 1.7 hours of reaction time. To optimize temperature, the engineer then fixes time at 1.7 hours (the apparent optimum) and performs five runs at different temperatures, say, 140, 150, 160, 170, and 180°F. The results of this set of runs are plotted in Figure 14.8. Maximum yield occurs at about 155°F. Therefore, we would conclude that running

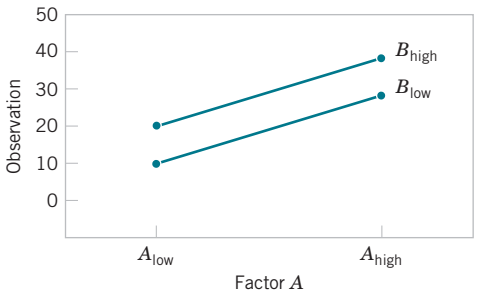


FIGURE 14.3
Factorial experiment, no interaction.

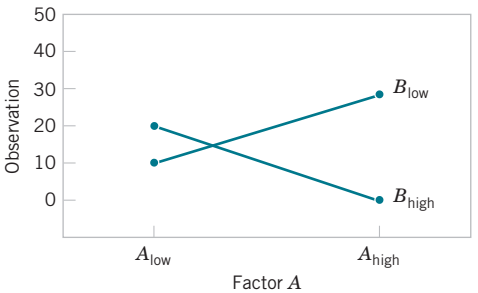


FIGURE 14.4
Factorial experiment, with interaction.

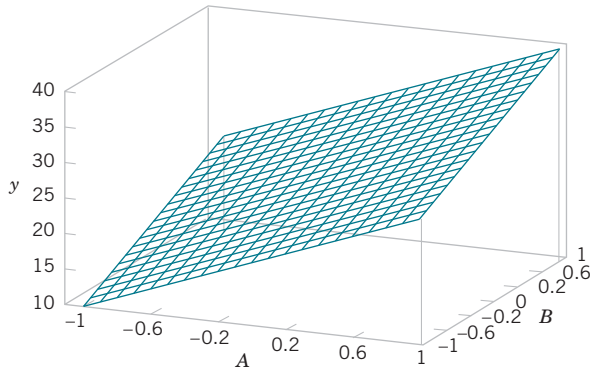


FIGURE 14.5

Three-dimensional surface plot of the data from Table 14.1, showing the main effects of the two factors *A* and *B*.

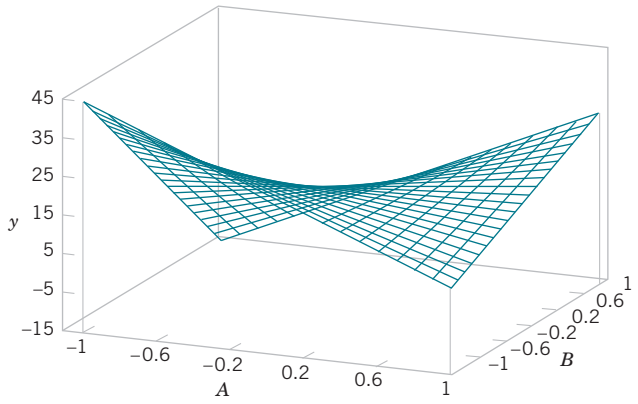


FIGURE 14.6

Three-dimensional surface plot of the data from Table 14.2, showing the effect of the *A* and *B* interaction.

the process at 155°F and 1.7 hours is the best set of operating conditions, resulting in yields of around 75%.

Figure 14.9 displays the contour plot of actual process yield as a function of temperature and time with the one-factor-at-a-time experiments superimposed on the contours. Clearly, this one-factor-at-a-time approach has failed dramatically here because the true optimum is at least 20 yield points higher and occurs at much lower reaction times and higher temperatures. The failure to discover the importance of the shorter reaction times is particularly important because this could have significant impact on production volume or capacity, production planning, manufacturing cost, and total productivity.

The one-factor-at-a-time approach has failed here because it cannot detect the interaction between temperature and time. Factorial experiments are the only way to detect interactions. Furthermore, the one-factor-at-a-time method is inefficient. It requires more experimentation than a factorial, and, as we have just seen, there is no assurance that it will produce the correct results.

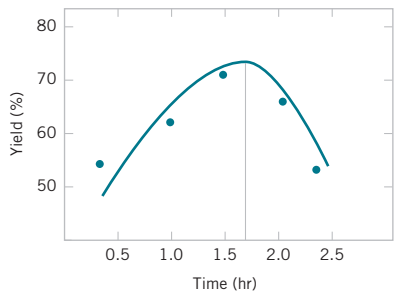


FIGURE 14.7

Yield versus reaction time with temperature constant at 155°F.

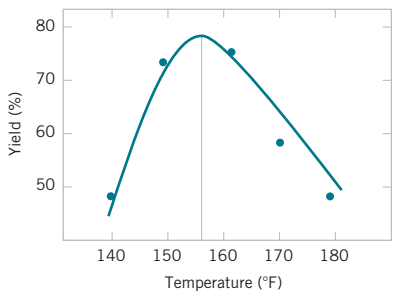


FIGURE 14.8

Yield versus temperature with reaction time constant at 1.7 hours.

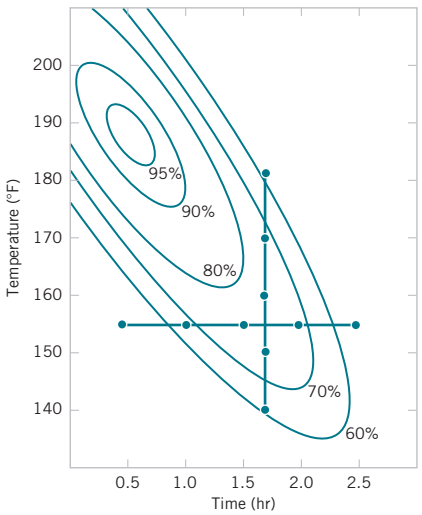


FIGURE 14.9

Optimization experiment using the one-factor-at-a-time method.

14.3 Two-Factor Factorial Experiments

The simplest type of factorial experiment involves only two factors, say A and B. There are a levels of factor A and b levels of factor B. This two-factor factorial is shown in Table 14.3. The experiment has n replicates, and each replicate contains all ab treatment combinations. The observation in the ijth cell for the kth replicate is denoted by y_{ijk}. In performing the experiment, the abn observations would be run in random order. Thus, like the single-factor experiment studied in Chapter 13, the two-factor factorial is a *completely randomized design*.

The observations may be described by the linear statistical model

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \\ k = 1, 2, \ldots, n \end{cases} \tag{14.1}$$

where μ is the overall mean effect, τ_i is the effect of the ith level of factor A, β_j is the effect of the jth level of factor B, $(\tau\beta)_{ij}$ is the effect of the interaction between A and B, and ϵ_{ijk} is a random error component having a normal distribution with mean 0 and variance σ^2. We are interested in testing the hypotheses of no main effect for factor A, no main effect for B, and no AB interaction effect. As with the single-factor experiments in Chapter 13, the analysis of variance (ANOVA) is used to test these hypotheses. Because the experiment has two factors, the test procedure is sometimes called a two-way analysis of variance.

14.3.1 Statistical Analysis

Suppose that the a levels of factor A and the b levels of factor B are specifically chosen by the experimenter, and inferences are confined to these levels only. In this model, it is customary to define the effects τ_i, β_j, and $(\tau\beta)_{ij}$ as deviations from the mean, so that $\sum_{i=1}^{a}\tau_i = 0$, $\sum_{j=1}^{b}\beta_j = 0$, $\sum_{i=1}^{a}(\tau\beta)_{ij} = 0$, and $\sum_{j=1}^{b}(\tau\beta)_{ij} = 0$.

The **analysis of variance** can be used to test hypotheses about the main factor effects of A and B and the AB interaction. To present the ANOVA, we need some symbols, some of which are illustrated in Table 14.3. Let $y_{i\cdot\cdot}$ denote the total of the observations taken at the ith level of factor A; $y_{\cdot j\cdot}$ denote the total of the observations taken at the jth level of factor B; $y_{ij\cdot}$ denote the total of the observations in the ijth cell of Table 14.3; and $y_{\cdots}$ denote the grand total of all the

TABLE 14.3 Data Arrangement for a Two-Factor Factorial Design

		Factor B					
		1	**2**	$\cdots$	***b***	**Totals**	**Averages**
	1	$y_{111}, y_{112}, \ldots, y_{11n}$	$y_{121}, y_{122}, \ldots, y_{12n}$		$y_{1b1}, y_{1b2}, \ldots, y_{1bn}$	$y_{1\cdot\cdot}$	$\bar{y}_{1\cdot\cdot}$
Factor *A*	2	$y_{211}, y_{212}, \ldots, y_{21n}$	$y_{221}, y_{222}, \ldots, y_{22n}$		$y_{2b1}, y_{2b2}, \ldots, y_{2bn}$	$y_{2\cdot\cdot}$	$\bar{y}_{2\cdot\cdot}$
	$\vdots$						
	a	$y_{a11}, y_{a12}, \ldots, y_{a1n}$	$y_{a21}, y_{a22}, \ldots, y_{a2n}$		$y_{ab1}, y_{ab2}, \ldots, y_{abn}$	$y_{a\cdot\cdot}$	$\bar{y}_{a\cdot\cdot}$
Totals		$y_{\cdot 1\cdot}$	$y_{\cdot 2\cdot}$		$y_{\cdot b\cdot}$	$y_{\cdots}$	
Averages		$\bar{y}_{\cdot 1\cdot}$	$\bar{y}_{\cdot 2\cdot}$		$\bar{y}_{\cdot b\cdot}$		$\bar{y}_{\cdots}$

observations. Define $\bar{y}_{i\cdot\cdot}, \bar{y}_{\cdot j\cdot}, \bar{y}_{ij\cdot}$, and $\bar{y}_{\cdots}$ as the corresponding row, column, cell, and grand averages. That is,

Notation for Totals and Means

$$y_{i\cdot\cdot} = \sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{i\cdot\cdot} = \frac{y_{i\cdot\cdot}}{bn} \qquad i = 1, 2, \ldots, a$$

$$y_{j\cdot\cdot} = \sum_{i=1}^{a}\sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{\cdot j\cdot} = \frac{y_{\cdot j\cdot}}{an} \qquad j = 1, 2, \ldots, b$$

$$y_{ij\cdot} = \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{ij\cdot} = \frac{y_{ij\cdot}}{n} \qquad i = 1, 2, \ldots, a, \quad j = 1, 2, \ldots, b$$

$$y_{\cdots} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{\cdots} = \frac{y_{\cdots}}{abn} \qquad i = 1, 2, \ldots, a$$

The hypotheses that we test are as follows:

$$\begin{array}{lll} H_0\colon \tau_1 = \tau_2 = \ldots = \tau_a = 0 & H_1\colon \text{at least one } \tau_i \neq 0 & \text{(no main effect of factor } A\text{)} \\ H_0\colon \beta_1 = \beta_2 = \ldots = \beta_b = 0 & H_1\colon \text{at least one } \beta_j \neq 0 & \text{(no main effect of factor } B\text{)} \\ H_0\colon (\tau\beta)_{11} = (\tau\beta)_{12} = \ldots = (\tau\beta)_{ab} = 0 & H_1\colon \text{at least one } (\tau\beta)_{ij} \neq 0 & \text{(no interaction)} \end{array} \tag{14.2}$$

As before, the ANOVA tests these hypotheses by decomposing the total variability in the data into component parts and then comparing the various elements in this decomposition. Total variability is measured by the total sum of squares of the observations

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} (y_{ijk} - \bar{y}_{\cdots})^2$$

and the definition of the sum of squares decomposition follows.

ANOVA Sum of Squares Identity: Two Factors

The *sum of squares identity for a two-factor ANOVA* is

$$\begin{aligned} \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} (y_{ijk} - \bar{y}_{\cdots})^2 &= bn\sum_{i=1}^{a} (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots})^2 + an\sum_{j=1}^{b} (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})^2 \\ &\quad + n\sum_{i=1}^{a}\sum_{j=1}^{b} (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} (y_{ijk} - \bar{y}_{ij\cdot})^2 \end{aligned} \tag{14.3}$$

or symbolically,

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E \tag{14.4}$$

and degrees of freedom can be partitioned as

$$abn - 1 = a - 1 + b - 1 + (a - 1)(b - 1) + ab(n - 1)$$

or

$$df_{\text{Total}} = df_A + df_B + df_{AB} + df_{\text{Error}}$$

Equations 14.3 and 14.4 state that the total sum of squares SS_T is partitioned into a sum of squares for the row factor A (SS_A), a sum of squares for the column factor B (SS_B), a sum of squares for the interaction between A and B (SS_{AB}), and an error sum of squares (SS_E). There are $abn - 1$ total degrees of freedom. The main effects A and B have $a - 1$ and $b - 1$ degrees of freedom, and the interaction effect AB has $(a - 1)(b - 1)$ degrees of freedom. Within each of the ab cells in Table 14.3, there are $n - 1$ degrees of freedom between the n replicates, and observations in the same cell can differ only because of random error. Therefore, there are $ab(n - 1)$ degrees of freedom for error, so the degrees of freedom are partitioned according to

$$abn - 1 = (a - 1) + (b - 1) + (a - 1)(b - 1) + ab(n - 1)$$

If we divide each of the sum of squares on the right-hand side of Equation 14.4 by the corresponding number of degrees of freedom, we obtain the mean squares for A, B, the interaction, and error:

$$MS_A = \frac{SS_A}{a-1} \qquad MS_B = \frac{SS_B}{b-1} \qquad MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)} \qquad MS_E = \frac{SS_E}{ab(n-1)}$$

Assuming that factors A and B are fixed factors, it is not difficult to show that the expected values of these mean squares are

Expected Values of Mean Squares: Two Factors

$$E(MS_A) = E\left(\frac{SS_A}{a-1}\right) = \sigma^2 + \frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1} \qquad E(MS_B) = E\left(\frac{SS_B}{b-1}\right) = \sigma^2 + \frac{an\sum_{j=1}^{a}\beta_j^2}{b-1}$$

$$E(MS_{AB}) = E\left(\frac{SS_{AB}}{(a-1)(b-1)}\right) = \sigma^2 + \frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)}$$

$$E(MS_E) = E\left(\frac{SS_E}{ab(n-1)}\right) = \sigma^2$$

From examining these expected mean squares, it is clear that if the null hypotheses about main effects H_0: $\tau_i = 0$, H_0: $\beta_j = 0$, and the interaction hypothesis H_0: $(\tau\beta)_{ij} = 0$ are all true, all four mean squares are unbiased estimates of σ^2.

To test that the row factor effects are all equal to zero (H_0: $\tau_i = 0$), columns effects are equal to zero (H_0: $\beta_j = 0$), and interaction effects are equal to zero (H_0: $(\tau\beta)_{ij} = 0$), we use the ratios

***F* Tests for Effects**

$$F_0 = \frac{MS_A}{MS_E} \qquad F_0 = \frac{MS_B}{MS_E} \qquad F_0 = \frac{MS_{AB}}{MS_E}$$

respectively. The first statistic has an F distribution with $a - 1$ and $ab(n - 1)$ degrees of freedom if H_0: $\tau_i = 0$ is true. This null hypothesis is rejected at the α level of significance if $f_0 > f_{\alpha,a-1,ab(n-1)}$.

The second statistic has an F distribution with $b - 1$ and $ab(n - 1)$ degrees of freedom if H_0: $\beta_j = 0$ is true. This null hypothesis is rejected at the α level of significance if $f_0 > f_{\alpha,b-1,ab(n-1)}$.

TABLE 14.4 ANOVA Table for a Two-Factor Factorial, Fixed-Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
A treatments	SS_A	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$\frac{MS_A}{MS_E}$
B treatments	SS_B	$b-1$	$MS_B = \frac{SS_B}{b-1}$	$\frac{MS_B}{MS_E}$
Interaction	SS_{AB}	$(a-1)(b-1)$	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$\frac{MS_{AB}}{MS_E}$
Error	SS_E	$ab(n-1)$	$MS_E = \frac{SS_E}{ab(n-1)}$	
Total	SS_T	$abn-1$		

The last statistic to test the interaction effects has an F distribution with $(a-1)(b-1)$ and $ab(n-1)$ degrees of freedom if the null hypothesis H_0: $(\tau\beta)_{ij} = 0$. This hypothesis is rejected at the α level of significance if $f_0 > f_{\alpha,(a-1)(b-1),ab(n-1)}$.

It is usually best to conduct the test for interaction first and then to evaluate the main effects. If interaction is not significant, interpretation of the tests on the main effects is straightforward. However, as noted earlier in this section, when interaction is significant, the main effects of the factors involved in the interaction may not have much practical interpretative value. Knowledge of the interaction is usually more important than knowledge about the main effects.

Computational formulas for the sums of squares are easily obtained, but manual computation is tedious. We suggest relying on statistical software. The results are usually displayed in an ANOVA table, such as Table 14.4.

EXAMPLE 14.1 | Aircraft Primer Paint

Aircraft primer paints are applied to aluminum surfaces by two methods: dipping and spraying. The purpose of using the primer is to improve paint adhesion, and some parts can be primed using either application method. The process engineering group responsible for this operation is interested in learning whether three different primers differ in their adhesion properties. A factorial experiment was performed to investigate the effect of paint primer type and application method on paint adhesion. For each combination of primer type and application method, three specimens were painted, then a finish paint was applied and the adhesion force was measured. The data from the experiment are shown in Table 14.5. The circled numbers in the cells are the cell totals y_{ij}. The sums of squares required to perform the ANOVA are computed as follows:

$$SS_T = 10.72$$
$$SS_{\text{types}} = 4.58$$
$$SS_{\text{methods}} = 4.91$$
$$SS_{\text{interaction}} = 0.24$$

TABLE 14.5 Adhesion Force Data

Primer Type	Dipping	Spraying	$y_{i\cdot\cdot}$
1	4.0, 4.5, 4.3	5.4, 4.9, 5.6	28.7
2	5.6, 4.9, 5.4	5.8, 6.1, 6.3	34.1
3	3.8, 3.7, 4.0	5.5, 5.0, 5.0	27.0
$y_{\cdot j\cdot}$	40.2	49.6	$89.8 = y_{\cdots}$

and

$$SS_E = SS_T - SS_{\text{types}} - SS_{\text{methods}} - SS_{\text{interaction}}$$
$$= 10.72 - 4.58 - 4.91 - 0.24 = 0.99$$

The ANOVA is summarized in Table 14.6. The experimenter has decided to use $\alpha = 0.05$. Because $f_{0.05,2,12} = 3.89$ and $f_{0.05,1,12} = 4.75$, we conclude that the main effects of

primer type and application method affect adhesion force. Furthermore, because $1.5 < f_{0.05,2,12}$, there is no indication of interaction between these factors. The last column of Table 14.6 shows the *P*-value for each *F*-ratio. Notice that the *P*-values for the two test statistics for the main effects are considerably less than 0.05, and the *P*-value for the test statistic for the interaction is more than 0.05.

Practical Interpretation: See a graph of the cell adhesion force averages $\{\bar{y}_{ij\cdot}\}$ versus levels of primer type for each application method in Figure 14.10. The no-interaction conclusion is obvious in this graph because the two lines are nearly parallel. Furthermore, because a large response indicates greater adhesion force, we conclude that spraying is the best application method and that primer type 2 is most effective.

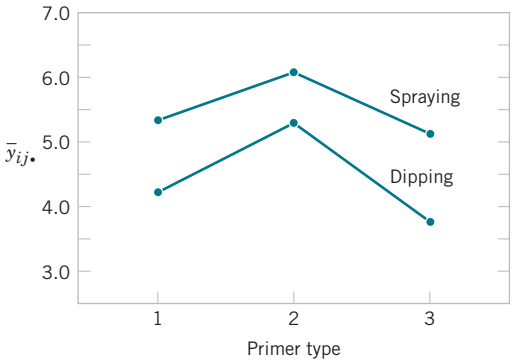


FIGURE 14.10
Graph of average adhesion force versus primer types for both application methods.

TABLE 14.6 ANOVA for Aircraft Primer Paint Experiment

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	*P*-Value
Primer types	4.58	2	2.29	27.86	2.7 × E-5
Application methods	4.91	1	4.91	59.70	4.7 × E-6
Interaction	0.24	2	0.12	1.47	0.2621
Error	0.99	12	0.08		
Total	10.72	17			

Tests on Individual Means When both factors are fixed, comparisons between the individual means of either factor may be made using any multiple comparison technique such as Fisher's LSD method (described in Chapter 13). When there is no interaction, these comparisons may be made using either the row averages $\bar{y}_{i\cdot\cdot}$ or the column averages $\bar{y}_{\cdot j\cdot}$. However, when interaction is significant, comparisons between the means of one factor (say, *A*) may be obscured by the *AB* interaction. In this case, we could apply a procedure such as Fisher's LSD method to the means of factor *A*, with factor *B* set at a particular level.

14.3.2 Model Adequacy Checking

Just as in the single-factor experiments discussed in Chapter 13, the residuals from a factorial experiment play an important role in assessing *model adequacy*. The residuals from a two-factor factorial are

$$e_{ijk} = y_{ijk} - \bar{y}_{ij\cdot}$$

That is, the residuals are just the difference between the observations and the corresponding cell averages.

Refer to Table 14.7 for the residuals for the aircraft primer paint data in Example 14.1. See Figure 14.11(a) for the normal probability plot of these residuals. This plot has tails that do not fall exactly along a straight line passing through the center of the plot, indicating some potential

TABLE 14.7 **Residuals for the Aircraft Primer Paint Experiment in Example 14.1**

	Application Method					
Primer Type	Dipping			Spraying		
1	−0.27,	0.23,	0.03	0.10,	−0.40,	0.30
2	0.30,	−0.40,	0.10	−0.27,	0.03,	0.23
3	−0.03,	−0.13,	0.17	0.33,	−0.17,	−0.17

problems with the normality assumption, but the deviation from normality does not appear severe. Figures 14.11(b) and 14.11(c) plot the residuals versus the levels of primer types and application methods, respectively. There is some indication that primer type 3 results in slightly lower variability in adhesion force than the other two primers. The graph of residuals versus fitted values in Figure 14.11(d) does not reveal any unusual or diagnostic pattern.

14.3.3 One Observation per Cell

In some cases involving a two-factor factorial experiment, we may have only one replicate—that is, only one observation per cell. In this situation, there are exactly as many parameters in the

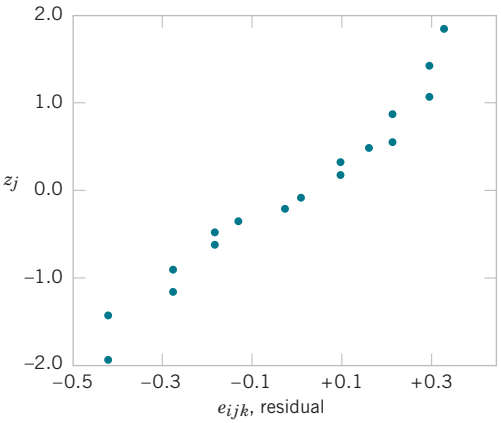


FIGURE 14.11(a)

Normal probability plot of the residuals from the aircraft primer paint experiment in Example 14.1.

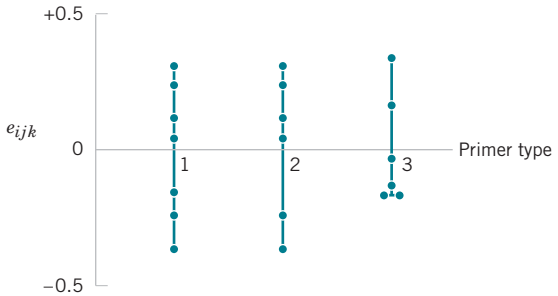


FIGURE 14.11(b)

Plot of residuals from the aircraft primer paint experiment versus primer type.

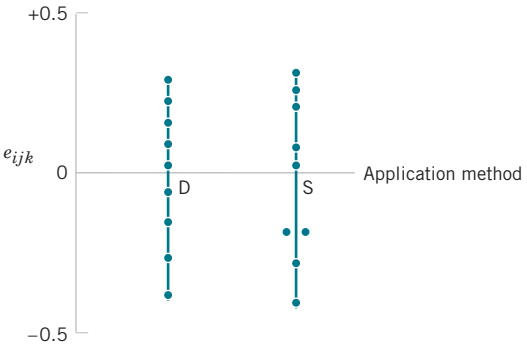


FIGURE 14.11(c)

Plot of residuals from the aircraft primer paint experiment versus application method.

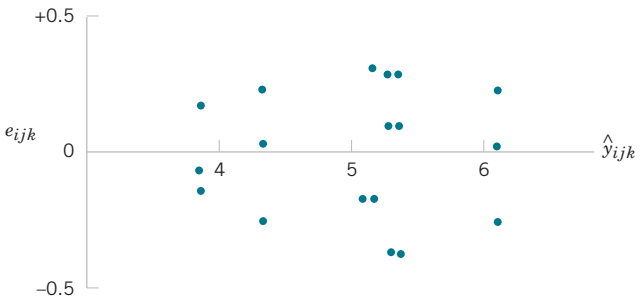


FIGURE 14.11(d)

Plot of residuals from the aircraft primer paint experiment versus predicted values $\hat{y}_{ijk}$.

analysis of variance model as observations, and the error degrees of freedom are zero. Thus, we cannot test hypotheses about the main effects and interactions unless we make some additional assumptions. One possible assumption is to assume that the interaction effect is negligible and use the interaction mean square as an error mean square. Thus, the analysis is equivalent to the analysis used in the randomized block design. This no-interaction assumption can be dangerous, and the experimenter should carefully examine the data and the residuals for indications of whether or not interaction is present.

14.4 General Factorial Experiments

Many experiments involve more than two factors. In this section, we introduce the case in which there are a levels of factor A, b levels of factor B, c levels of factor C, and so on, arranged in a factorial experiment. In general, there are $a \times b \times c \cdots \times n$ total observations if there are n replicates of the complete experiment.

For example, consider the *three-factor-factorial experiment*, with underlying model

$$Y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \epsilon_{ijkl} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \\ k = 1, 2, \ldots, c \\ l = 1, 2, \ldots, n \end{cases} \tag{14.5}$$

Notice that the model contains three main effects, three two-factor interactions, a three-factor interaction, and an error term. The analysis of variance is shown in Table 14.8. Note that there

TABLE 14.8 Analysis of Variance Table for the Three-Factor Fixed Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Expected Mean Square	F_0
A	SS_A	$a-1$	MS_A	$\sigma^2 + \frac{bcn\sum \tau_i^2}{a-1}$	$\frac{MS_A}{MS_E}$
B	SS_B	$b-1$	MS_B	$\sigma^2 + \frac{acn\sum \beta_j^2}{b-1}$	$\frac{MS_B}{MS_E}$
C	SS_C	$c-1$	MS_C	$\sigma^2 + \frac{abn\sum \gamma_k^2}{c-1}$	$\frac{MS_C}{MS_E}$
AB	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	$\sigma^2 + \frac{cn\sum\sum (\tau\beta)_{ij}^2}{(a-1)(b-1)}$	$\frac{MS_{AB}}{MS_E}$
AC	SS_{AC}	$(a-1)(c-1)$	MS_{AC}	$\sigma^2 + \frac{bn\sum\sum (\tau\gamma)_{ik}^2}{(a-1)(c-1)}$	$\frac{MS_{AC}}{MS_E}$
BC	SS_{BC}	$(b-1)(c-1)$	MS_{BC}	$\sigma^2 + \frac{cn\sum\sum (\beta\gamma)_{jk}^2}{(b-1)(c-1)}$	$\frac{MS_{BC}}{MS_E}$
ABC	SS_{ABC}	$(a-1)(b-1)(c-1)$	MS_{ABC}	$\sigma^2 + \frac{n\sum\sum\sum (\tau\beta\gamma)_{ijk}^2}{(a-1)(b-1)(c-1)}$	$\frac{MS_{ABC}}{MS_E}$
Error	SS_E	$abc(n-1)$	MS_E	σ^2	
Total	SS_T	$abcn-1$			

must be at least two replicates ($n \geq 2$) to compute an error sum of squares. The F-test on main effects and interactions follows directly from the expected mean squares. These ratios follow F-distributions under the respective null hypotheses.

EXAMPLE 14.2 | Surface Roughness

A mechanical engineer is studying the surface roughness of a part produced in a metal-cutting operation. Three factors, feed rate (A), depth of cut (B), and tool angle (C), are of interest. All three factors have been assigned two levels, and two replicates of a factorial design are run. The coded data are in Table 14.9.

TABLE 14.9 Coded Surface Roughness Data

	Depth of Cut (B)			
	0.025 inch		0.040 inch	
	Tool Angle (C)		Tool Angle (C)	
Feed Rate (A)	15°	25°	15°	25°
20 inches per minute	9	11	9	10
	7	10	11	8
30 inches per minute	10	10	12	16
	12	13	15	14

The ANOVA is summarized in Table 14.10. Because manual ANOVA computations are tedious for three-factor experiments, we have used computer software for the solution of this problem. The F-ratios for all three main effects and the interactions are formed by dividing the mean square for the effect of interest by the error mean square. Because the experimenter has selected $\alpha = 0.05$, the critical value for each of these F-ratios is $f_{0.05,1,8} = 5.32$. Alternately, we could use the P-value approach. The P-values for all the test statistics are shown in the last column of Table 14.10. Inspection of these P-values is revealing. There is a strong main effect of feed rate because the F-ratio is well into the critical region. However, there is some indication of an effect due to the depth of cut because $P = 0.0710$ is not much greater than $\alpha = 0.05$. The next largest effect is the AB or feed rate × depth of cut interaction. Most likely, both feed rate and depth of cut are important process variables.

Practical Interpretation: Further experiments could study the important factors in more detail to improve the surface roughness.

TABLE 14.10 ANOVA

ANOVA

Factor	Levels	Values	
Feed	2	20	30
Depth	2	0.025	0.040
Angle	2	15	25

Analysis of Variance for Roughness

Source	DF	SS	MS	F	P
Feed	1	45.563	45.563	18.69	0.003
Depth	1	10.563	10.563	4.33	0.071
Angle	1	3.063	3.063	1.26	0.295
Feed*Depth	1	7.563	7.563	3.10	0.116
Feed*Angle	1	0.062	0.062	0.03	0.877
Depth*Angle	1	1.563	1.563	0.64	0.446
Feed*Depth*Angle	1	5.062	5.062	2.08	0.188
Error	8	19.500	2.437		
Total	15	92.938			

Obviously, factorial experiments with three or more factors can require many runs, particularly if some of the factors have several (more than two) levels. This point of view leads us to the class of factorial designs (considered in a later section) with all factors at two levels. These

designs are easy to set up and analyze, and they may be used as the basis of many other useful experimental designs.

14.5 2^k Factorial Designs

Factorial designs are frequently used in experiments involving several factors where it is necessary to study the joint effect of the factors on a response. However, several special cases of the general factorial design are important because they are widely employed in research work and because they form the basis of other designs of considerable practical value.

The most important of these special cases is that of k factors, each at only two levels. These levels may be quantitative, such as two values of temperature, pressure, or time; or they may be qualitative, such as two machines, two operators, the "high" and "low" levels of a factor, or perhaps the presence and absence of a factor. A complete replicate of such a design requires $2 \times 2 \times \cdots \times 2 = 2^k$ observations and is called a **2^k factorial design**.

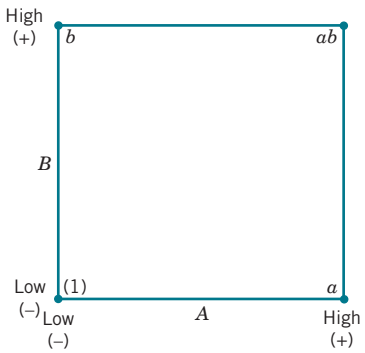
The 2^k design is particularly useful in the early stages of experimental work when many factors are likely to be investigated. It provides the smallest number of runs for which k factors can be studied in a complete factorial design. Because each factor has only two levels, we must assume that the response is approximately linear over the range of the factor levels chosen.

14.5.1 2^2 Design

The simplest type of 2^k design is the 2^2—that is, two factors A and B, each at two levels. We usually think of these levels as the factor's low and high levels. The 2^2 design is shown in Figure 14.12. Note that the design can be represented geometrically as a square with the $2^2 = 4$ runs, or treatment combinations, forming the corners of the square. In the 2^2 design, it is customary to denote the low and high levels of the factors A and B by the signs $-$ and $+$, respectively. This is sometimes called the **geometric notation** for the design.

A special notation is used to label the treatment combinations. In general, a treatment combination is represented by a series of lowercase letters. If a letter is present, the corresponding factor is run at the high level in that treatment combination; if it is absent, the factor is run at its low level. For example, treatment combination a indicates that factor A is at the high level and factor B is at the low level. The treatment combination with both factors at the low level is represented by (1). This notation is used throughout the 2^k design series. For example, the treatment combination in a 2^4 with A and C at the high level and B and D at the low level is denoted by ac.

The effects of interest in the 2^2 design are the main effects A and B and the two-factor interaction AB. Let the symbols (1), a, b, and ab also represent the totals of all n observations taken at these design points. It is easy to estimate the effects of these factors. To estimate the main effect of A, we would average the observations on the right side of the square in Figure 14.12 where



Treatment	A	B
(1)	–	–
a	+	–
b	–	+
ab	+	+

FIGURE 14.12
The 2^2 factorial design.

A is at the high level, and subtract from this the average of the observations on the left side of the square where A is at the low level. Similarly, the main effect of B is found by averaging the observations on the top of the square, where B is at the high level, and subtracting the average of the observations on the bottom of the square, where B is at the low level.

Main Effects of Factors: 2^2 Design

$$A = \bar{y}_{A+} - \bar{y}_{A-} = \frac{a + ab}{2n} - \frac{b + (1)}{2n} = \frac{1}{2n}[a + ab - b - (1)] \tag{14.6}$$

$$B = \bar{y}_{B+} - \bar{y}_{B-} = \frac{b + ab}{2n} - \frac{a + (1)}{2n} = \frac{1}{2n}[b + ab - a - (1)] \tag{14.7}$$

Finally, the AB interaction is estimated by taking the difference in the diagonal averages in Figure 14.12, or

Interaction Effect AB: 2^2 Design

$$AB = \frac{ab + (1)}{2n} - \frac{a + b}{2n} = \frac{1}{2n}[ab + (1) - a - b] \tag{14.8}$$

The quantities in brackets in Equations 14.6, 14.7, and 14.8 are called **contrasts**. For example, the A contrast is

$$\text{Contrast}_A = a + ab - b - (1)$$

In these equations, the contrast coefficients are always either +1 or −1. A table of plus and minus signs, such as Table 14.11, can be used to determine the sign on each treatment combination for a particular contrast. The column headings for Table 14.11 are the main effects A and B, the AB interaction, and I, which represents the total. The row headings are the treatment combinations. Note that the signs in the AB column are the product of signs from columns A and B. To generate a contrast from this table, multiply the signs in the appropriate column of Table 14.11 by the treatment combinations listed in the rows and add. For example, $\text{contrast}_{AB} = [(1)] + [-a] + [-b] + [ab] = ab + (1) - a - b$.

TABLE 14.11 **Signs for Effects in the 2^2 Design**

Treatment Combination	Factorial Effect			
	I	A	B	AB
(1)	+	−	−	+
a	+	+	−	−
b	+	−	+	−
ab	+	+	+	+

Contrasts are used in calculating both the effect estimates and the sums of squares for A, B, and the AB interaction. For any 2^k design with n replicates, the effect estimates are computed from

Relationship Between a Contrast and an Effect

$$\text{Effect} = \frac{\text{Contrast}}{n2^{k-1}} \tag{14.9}$$

and the sum of squares for any effect is

Sum of Squares for an Effect

$$SS = \frac{(\text{Contrast})^2}{n2^k} \tag{14.10}$$

One degree of freedom is associated with each effect (two levels minus one) so that the mean square of each effect equals the sum of squares. The analysis of variance is completed by computing the total sum of squares SS_T (with $4n - 1$ degrees of freedom) as usual, and obtaining the error sum of squares SS_E (with $4(n - 1)$ degrees of freedom) by subtraction.

EXAMPLE 14.3 | Epitaxial Process

An article in the *AT&T Technical Journal* (March/April 1986, Vol. 65, pp. 39–50) describes the application of two-level factorial designs to integrated circuit manufacturing. A basic processing step in this industry is to grow an epitaxial layer on polished silicon wafers. The wafers are mounted on a susceptor and positioned inside a bell jar. Chemical vapors are introduced through nozzles near the top of the jar. The susceptor is rotated, and heat is applied. These conditions are maintained until the epitaxial layer is thick enough.

Refer to Table 14.12 for the results of a 2^2 factorial design with $n = 4$ replicates using the factors A = deposition time and B = arsenic flow rate. The two levels of deposition time are $-$ = short and $+$ = long, and the two levels of arsenic flow rate are $-$ = 55% and $+$ = 59%. The response variable is epitaxial layer thickness (μm). We may find the estimates of the effects using Equations 14.6, 14.7, and 14.8 as follows:

$$A = \frac{1}{2n}[a + ab - b - (1)]$$

$$= \frac{1}{2(4)}[59.299 + 59.156 - 55.686 - 56.081] = 0.836$$

$$B = \frac{1}{2n}[b + ab - a - (1)]$$

$$= \frac{1}{2(4)}[55.686 + 59.156 - 59.299 - 56.081] = -0.067$$

$$AB = \frac{1}{2n}[ab + (1) - a - b]$$

$$= \frac{1}{2(4)}[59.156 + 56.081 - 59.299 - 55.686] = 0.032$$

TABLE 14.12 The 2^2 Design for the Epitaxial Process Experiment

Treatment Combination	Design Factors			Thickness (μm)				Thickness (μm)	
	A	B	AB	Thickness (μm)				Total	Average
(1)	−	−	+	14.037	14.165	13.972	13.907	56.081	14.020
a	+	−	−	14.821	14.757	14.843	14.878	59.299	14.825
b	−	+	−	13.880	13.860	14.032	13.914	55.686	13.922
ab	+	+	+	14.888	14.921	14.415	14.932	59.156	14.789

The numerical estimates of the effects indicate that the effect of deposition time is large and has a positive direction (increasing deposition time increases thickness) because changing deposition time from low to high changes the mean epitaxial layer thickness by 0.836 μm. The effects of arsenic flow rate (B) and the AB interaction appear small.

The importance of these effects may be confirmed with the analysis of variance. The sums of squares for A, B, and AB are computed as follows:

$$SS_A = \frac{[a + ab - b - (1)]^2}{16} = \frac{[6.688]^2}{16} = 2.7956$$

$$SS_B = \frac{[b + ab - a - (1)]^2}{16} = \frac{[-0.538]^2}{16} = 0.0181$$

$$SS_{AB} = \frac{[ab + (1) - a - b]^2}{16} = \frac{[0.252]^2}{16} = 0.0040$$

$$SS_T = 14.037^2 + \cdots + 14.932^2 - \frac{(56.081 + \cdots + 59.156)^2}{16}$$
$$= 3.0672$$

Practical Interpretation: The analysis of variance is summarized in the bottom half of Table 14.13 and confirms our conclusions obtained by examining the magnitude and direction of the effects. Deposition time is the only factor that significantly affects epitaxial layer thickness, and from the direction of the effect estimates, we know that longer deposition times lead to thicker epitaxial layers. The top half of Table 14.13 is discussed next.

TABLE 14.13 Analysis for the Epitaxial Process Experiment

Term	Effect	Coefficient	SE Coefficient	t	P-Value
Constant	14.3889	0.03605	399.17	0.000	
A	0.8360	0.4180	0.03605	11.60	0.00
B	−0.0672	−0.0336	0.03605	−0.93	0.38
AB	0.0315	0.0157	0.03605	0.44	0.67

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-Value
A (deposition time)	2.7956	1	2.7956	134.40	0.00
B (arsenic flow)	0.0181	1	0.0181	0.87	0.38
AB	0.0040	1	0.0040	0.19	0.67
Error	0.2495	12	0.0208		
Total	3.0672	15			

Models and Residual Analysis It is easy to obtain a model for the response and residuals from a 2^k design by fitting a **regression model** to the data. For the epitaxial process experiment, the regression model is

$$Y = \beta_0 + \beta_1 x_1 + \epsilon$$

because the only active variable is deposition time, which is represented by a coded variable x_1. The low and high levels of deposition time are assigned values $x_1 = -1$ and $x_1 = +1$, respectively. The least squares fitted model is

$$\hat{y} = 14.389 + \left(\frac{0.836}{2}\right) x_1$$

where the intercept $\hat{\beta}_0$ is the grand average of all 16 observations ($\bar{y}$) and the slope $\hat{\beta}_1$ is one-half the effect estimate for deposition time. The regression coefficient is one-half the effect estimate because regression coefficients measure the effect of a unit change in x_1 on the mean of Y, and the effect estimate is based on a two-unit change from −1 to +1.

A coefficient relates a factor to the response and, similar to regression analysis, interest centers on whether or not a coefficient estimate is significantly different from zero. Each effect estimate in Equations 14.6 through 14.8 is the difference between two averages (that we denote

in general as $\bar{y}_+ - \bar{y}_-$). In a 2^k experiment with n replicates, half the observations appear in each average so that there are $n2^{k-1}$ observations in each. The associated coefficient estimate, say $\hat{\beta}$, equals half the associated effect estimate so that

Relationship Between a Coefficient and an Effect

$$\hat{\beta} = \frac{\text{effect}}{2} = \frac{\bar{y}_+ - \bar{y}_-}{2} \tag{14.11}$$

The standard error of $\hat{\beta}$ is obtained from the standard error of the difference between two averages. Therefore,

Standard Error of a Coefficient

$$\text{Standard error } \hat{\beta} = \frac{\hat{\sigma}}{2}\sqrt{\frac{1}{n2^{k-1}} + \frac{1}{n2^{k-1}}} = \hat{\sigma}\sqrt{\frac{1}{n2^k}} \tag{14.12}$$

where $\hat{\sigma}$ is estimated from the square root of the error mean square. A t-test for a coefficient can also be used to test the significance of an effect. The t-statistic to test H_0: $\beta = 0$ in a 2^k experiment is

***t*-statistic to Test That a Coefficient Is Zero**

$$t = \frac{\hat{\beta}}{\text{Standard error } \hat{\beta}} = \frac{(\bar{y}_+ - \bar{y}_-)/2}{\hat{\sigma}\sqrt{\frac{1}{n2^k}}} \tag{14.13}$$

with degrees of freedom equal to those associated with error mean square. This statistic is similar to a two-sample t-test, but σ is estimated from the root of the error mean square. The estimate $\hat{\sigma}$ accounts for the multiple treatments in an experiment and generally differs from the estimate used in a two-sample t-test.

Some algebra can be used to show that for a 2^k experiment, the square of the t-statistic for the coefficient test equals the F-statistic used for the effect test in the analysis of variance. Furthermore, the square of a random variable with a t distribution with d degrees of freedom has an F distribution with 1 numerator and d denominator degrees of freedom. Thus, the test that compares the absolute value of the t-statistic to the t distribution is equivalent to the ANOVA F-test, with the exact same P-value, and either method may be used to test an effect.

For example, for the epitaxial process experiment in Example 14.3, the effect of A is 0.836. Therefore, the coefficient for A is $0.836/2 = 0.418$. Furthermore, $\hat{\sigma}^2 = 0.0208$ from the error mean square in the ANOVA table. Therefore, the standard error of a coefficient is $\sqrt{0.0208/[4(2^2)]} = 0.03605$, and the t-statistic for factor A is $0.418/0.03605 = 11.60$. The square of the t-statistic is $11.596^2 = 134.47$, and this equals the F-statistic for factor A in the ANOVA. The upper half of Table 14.13 shows the results for the other coefficients. Notice that the P-values obtained from the t-tests equal those in the ANOVA table. The analysis of a 2^k design through coefficient estimates and t-tests is similar to the approach used in regression analysis. Consequently, it might be easier to interpret results from this perspective. Computer software often generates output in this format.

The least squares fitted model can also be used to obtain the predicted values at the four points that form the corners of the square in the design. For example, consider the point with low deposition time ($x_1 = -1$) and low arsenic flow rate. The predicted value is

$$\hat{y} = 14.389 + \left(\frac{0.836}{2}\right)(-1) = 13.971 \ \mu\text{m}$$

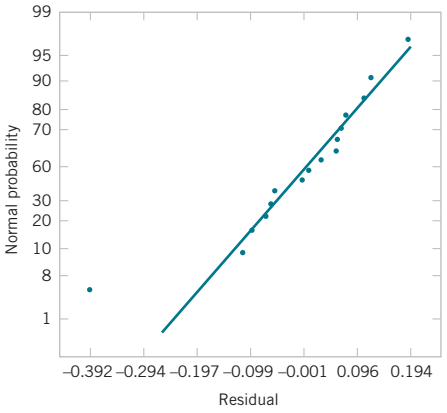


FIGURE 14.13
Normal probability plot of residuals for the epitaxial process experiment.

and the residuals for the four runs at that design point are

$$e_1 = 14.037 - 13.971 = 0.066$$
$$e_2 = 14.165 - 13.971 = 0.194$$
$$e_3 = 13.972 - 13.971 = 0.001$$
$$e_4 = 13.907 - 13.971 = -0.064$$

The remaining predicted values and residuals at the other three design points are calculated in a similar manner.

A normal probability plot of these residuals is shown in Figure 14.13. This plot indicates that one residual $e_{15} = -0.392$ is an outlier. Examining the four runs with high deposition time and high arsenic flow rate reveals that observation $y_{15} = 14.415$ is considerably smaller than the other three observations at that treatment combination. This adds some additional evidence to the tentative conclusion that observation 15 is an outlier. Another possibility is that some process variables affect the *variability* in epitaxial layer thickness. If we could discover which variables produce this effect, we could perhaps adjust these variables to levels that would minimize the variability in epitaxial layer thickness. Figures 14.14 and 14.15 are plots of residuals versus deposition time and arsenic flow rate, respectively. Except for the unusually large residual associated with y_{15}, there is no strong evidence that either deposition time or arsenic flow rate influences the variability in epitaxial layer thickness.

Figure 14.16 shows the standard deviation of epitaxial layer thickness at all four runs in the 2^2 design. These standard deviations were calculated using the data in Table 14.12. Notice that the standard deviation of the four observations with A and B at the high level is considerably larger than the standard deviations at any of the other three design points. Most of this difference is attributable to the unusually low thickness measurement associated with y_{15}. The standard deviation of the four observations with A and B at the low level is also somewhat larger than the standard deviations at the remaining two runs. This could indicate that other process variables not included in this experiment may affect the variability in epitaxial layer thickness. Another experiment to study this possibility, involving other process variables, could be designed and

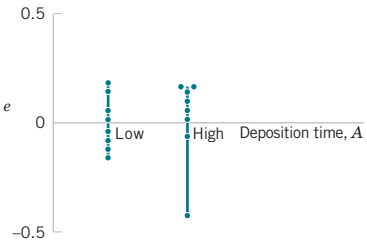


FIGURE 14.14
Plot of residuals versus deposition time.

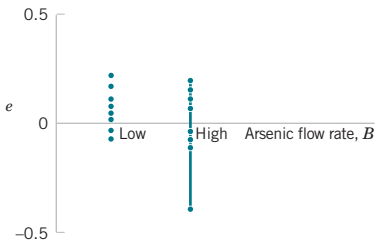


FIGURE 14.15
Plot of residuals versus arsenic flow rate.

conducted. (The original paper in the *AT&T Technical Journal* shows that two additional factors not considered in this example affect process variability.)

14.5.2 2^k Design for $k \geq 3$ Factors

The methods presented in the previous section for factorial designs with $k = 2$ factors each at two levels can be easily extended to more than two factors. For example, consider $k = 3$ factors, each at two levels. This design is a 2^3 factorial design, and it has eight runs or treatment combinations. Geometrically, the design is a cube as shown in Figure 14.17(a) with the eight runs forming the corners of the cube. Figure 14.17(b) lists the eight runs in a table with each row representing one of the runs and the − and + settings indicating the low and high levels for each of the three factors. This table is sometimes called the **design matrix**. This design allows three main effects (A, B, and C) to be estimated along with three two-factor interactions (AB, AC, and BC) and a three-factor interaction (ABC).

The main effects can easily be estimated. Remember that the lowercase letters (1), a, b, ab, c, ac, bc, and abc represent the total of all n replicates at each of the eight runs in the design. As in Figure 14.18(a), the main effect of A can be estimated by averaging the four treatment combinations on the right-hand side of the cube where A is at the high level and by subtracting from this quantity the average of the four treatment combinations on the left-hand side of the cube where A is at the low level. This gives

$$A = \bar{y}_{A+} - \bar{y}_{A-} = \frac{a + ab + ac + abc}{4n} - \frac{(1) + b + c + bc}{4n}$$

This equation can be rearranged as follows, and in a similar manner we determine the B and C effects from differences in averages [Figure 14.18(a)]

Main Effects of Factors: 2^3 Design

$$A = \bar{y}_{A+} - \bar{y}_{A-} = \frac{1}{4n}[a + ab + ac + abc - (1) - b - c - bc]$$

$$B = \bar{y}_{B+} - \bar{y}_{B-} = \frac{1}{4n}[b + ab + bc + abc - (1) - a - c - ac]$$

$$C = \bar{y}_{C+} - \bar{y}_{C-} = \frac{1}{4n}[c + ac + bc + abc - (1) - a - b - ab]$$

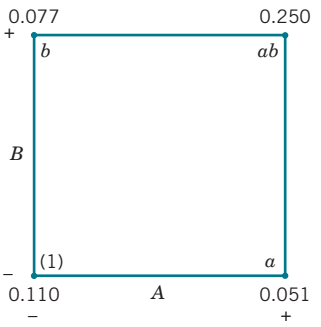
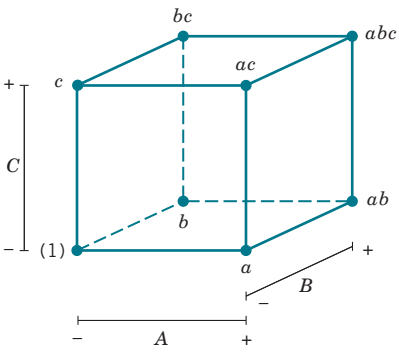


FIGURE 14.16
The standard deviation of epitaxial layer thickness at the four runs in the 2^2 design.



(a) Geometric view

Run	A	B	C
1	−	−	−
2	+	−	−
3	−	+	−
4	+	+	−
5	−	−	+
6	+	−	+
7	−	+	+
8	+	+	+

(b) 2^3 design matrix

FIGURE 14.17
The 2^3 design.

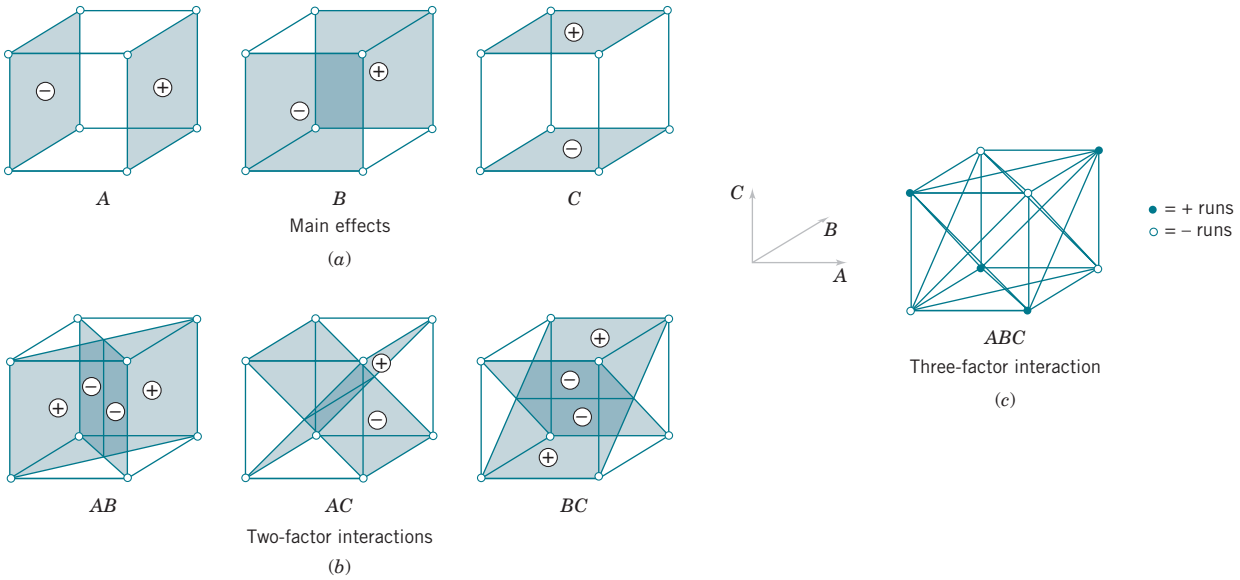


FIGURE 14.18

Geometric presentation of contrasts corresponding to the main effects and interaction in the 2^3 design. (a) Main effects. (b) Two-factor interactions. (c) Three-factor interaction.

The two-factor interaction effects may be computed easily. A measure of the AB interaction is the difference between the average A effects at the two levels of B. Symbolically,

B	**Average *A* Effect**
High (+)	$\dfrac{[(abc - bc) + (ab - b)]}{2n}$
Low (−)	$\dfrac{\{(ac - c) + [a - (1)]\}}{2n}$
Difference	$\dfrac{[abc - bc + ab - b - ac + c - a + (1)]}{2n}$

By convention, the AB interaction is one-half of this difference, and the AC and BC interaction effects are obtained in a similar manner.

Two-Factor Interaction Effects: 2^3 Design

$$AB = \frac{1}{4n}[abc - bc + ab - b - ac + c - a + (1)]$$

$$AC = \frac{1}{4n}[(1) - a + b - ab - c + ac - bc + abc]$$

$$BC = \frac{1}{4n}[(1) + a - b - ab - c - ac + bc + abc]$$

We could write the AB effect as follows:

$$AB = \frac{abc + ab + c + (1)}{4n} - \frac{bc + b + ac + a}{4n}$$

In this form, the AB interaction can easily be seen to be the difference in averages between runs on two diagonal planes in the cube in Figure 14.18(b).

The ABC interaction is defined as the average difference between the AB interaction for the two different levels of C. Thus,

$$ABC = \frac{1}{4n}\{[abc - bc] - [ac - c] - [ab - b] + [a - (1)]\}$$

or

Three-Factor Interaction Effect: 2^3 Design

$$ABC = \frac{1}{4n}[abc - bc - ac + c - ab + b + a - (1)]$$

As before, we can think of the ABC interaction as the difference in two averages. If the runs in the two averages are isolated, they define the vertices of the two tetrahedra that comprise the cube in Figure 14.18(c).

In the equations for the effects, the quantities in brackets are contrasts in the treatment combinations. A table of plus and minus signs can be developed from the contrasts and is shown in Table 14.14. Signs for the main effects are determined directly from the test matrix in Figure 14.17(b). Once the signs for the main effect columns have been established, the signs for the remaining columns can be obtained by multiplying the appropriate main effect row by row. For example, the signs in the AB column are the products of the A and B column signs in each row. The contrast for any effect can easily be obtained from this table.

Table 14.14 has several interesting properties:

1. Except for the identity column I, each column has an equal number of plus and minus signs.
2. The sum of products of signs in any two columns is zero; that is, the columns in the table are *orthogonal*.
3. Multiplying any column by column I leaves the column unchanged; that is, I is an *identity element*.
4. The product of any two columns yields a column in the table, for example $A \times B = AB$, and $AB \times ABC = A^2B^2C = C$ because any column multiplied by itself is the identity column.

TABLE 14.14 Algebraic Signs for Calculating Effects in the 2^3 Design

Treatment Combination	Factorial Effect							
	I	A	B	AB	C	AC	BC	ABC
(1)	+	−	−	+	−	+	+	−
a	+	+	−	−	−	−	+	+
b	+	−	+	−	−	+	−	+
ab	+	+	+	+	−	−	−	−
c	+	−	−	+	+	−	−	+
ac	+	+	−	−	+	+	−	−
bc	+	−	+	−	+	−	+	−
abc	+	+	+	+	+	+	+	+

The estimate of any main effect or interaction in a 2^k design is determined by multiplying the treatment combinations in the first column of the table by the signs in the corresponding main effect or interaction column, by adding the result to produce a contrast, and then by dividing the contrast by one-half the total number of runs in the experiment.

EXAMPLE 14.4 | Surface Roughness

Consider the surface roughness experiment originally described in Example 14.2. This is a 2^3 factorial design in the factors feed rate (A), depth of cut (B), and tool angle (C), with $n = 2$ replicates. Table 14.15 presents the observed surface roughness data.

The effect of A, for example, is

$$\begin{aligned} A &= \frac{1}{4n}\left[a + ab + ac + abc - (1) - b - c - bc\right] \\ &= \frac{1}{4(2)}\left[22 + 27 + 23 + 30 - 16 - 20 - 21 - 18\right] \\ &= \frac{1}{8}\left[27\right] = 3.375 \end{aligned}$$

and the sum of squares for A is found using Equation 14.10:

$$SS_A = \frac{(\text{contrast}_A)^2}{n2^k} = \frac{(27)^2}{2(8)} = 45.5625$$

It is easy to verify that the other effects are

$$\begin{aligned} B &= 1.625 \\ C &= 0.875 \\ AB &= 1.375 \\ AC &= 0.125 \\ BC &= -0.625 \\ ABC &= 1.125 \end{aligned}$$

Examining the magnitude of the effects clearly shows that feed rate (factor A) is dominant, followed by depth of cut (B) and the AB interaction, although the interaction effect is relatively small. The analysis, summarized in Table 14.16, confirms our interpretation of the effect estimates.

The output from the computer software for this experiment is shown in Table 14.17. The upper portion of the table displays the effect estimates and regression coefficients for each factorial effect. To illustrate, for the main effect of feed, the computer output reports $t = 4.32$ (with 8 degrees of freedom), and $t^2 = (4.32)^2 = 18.66$, which is approximately equal to the F-ratio for feed reported in Table 14.17 ($F = 18.69$). This F-ratio has one numerator and eight denominator degrees of freedom.

The lower panel of the computer output in Table 14.17 is an analysis of variance summary focusing on the types of terms in the model. A regression model approach is used in the presentation. You might find it helpful to review Section 12.2.2, particularly the material on the partial F-test. The row entitled "main effects" under source refers to the three main effects feed, depth, and angle, each having a single degree of freedom, giving the total 3 in the column headed "DF." The column headed "Seq SS" (an abbreviation for sequential sum of squares) reports how much the model sum of squares increases when each group of terms is added to a model that contains the terms listed *above* the groups. The first number in the "Seq SS" column presents the model sum of squares

TABLE 14.15 Surface Roughness Data

Treatment Combinations	Design Factors			Surface Roughness	Totals
	A	*B*	*C*		
(1)	−1	−1	−1	9, 7	16
a	1	−1	−1	10, 12	22
b	−1	1	−1	9, 11	20
ab	1	1	−1	12, 15	27
c	−1	−1	1	11, 10	21
ac	1	−1	1	10, 13	23
bc	−1	1	1	10, 8	18
abc	1	1	1	16, 14	30

TABLE 14.16 Analysis for the Surface Roughness Experiment

Term	Effect	Coefficient	SE Coefficient	t	P-Value
Constant		11.0625	0.3903	28.34	0.000
A	3.3750	1.6875	0.3903	4.32	0.003
B	1.6250	0.8125	0.3903	2.08	0.071
C	0.8750	0.4375	0.3903	1.12	0.295
AB	1.3750	0.6875	0.3903	1.76	0.116
AC	0.1250	0.0625	0.3903	0.16	0.877
BC	−0.6250	−0.3125	0.3903	−0.80	0.446
ABC	1.1250	0.5625	0.3903	1.44	0.188

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-Value
A	45.5625	1	45.5625	18.69	0.002
B	10.5625	1	10.5625	4.33	0.071
C	3.0625	1	3.0625	1.26	0.295
AB	7.5625	1	7.5625	3.10	0.116
AC	0.0625	1	0.0625	0.03	0.877
BC	1.5625	1	1.5625	0.64	0.446
ABC	5.0625	1	5.0625	2.08	0.188
Error	19.5000	8	2.4375		
Total	92.9375	15			

TABLE 14.17 Computer Analysis for the Surface Roughness Experiment

Estimated Effects and Coefficients for Roughness

Term	Effect	Coef	StDev Coef	T	P
Constant		11.0625	0.3903	28.34	0.000
Feed	3.3750	1.6875	0.3903	4.32	0.003
Depth	1.6250	0.8125	0.3903	2.08	0.071
Angle	0.8750	0.4375	0.3903	1.12	0.295
Feed*Depth	1.3750	0.6875	0.3903	1.76	0.116
Feed*Angle	0.1250	0.0625	0.3903	0.16	0.877
Depth*Angle	−0.6250	−0.3125	0.3903	−0.80	0.446
Feed*Depth*Angle	1.1250	0.5625	0.3903	1.44	0.188

Analysis of Variance for Roughness

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main effects	3	59.188	59.188	19.729	8.09	0.008
2-Way interactions	3	9.187	9.187	3.062	1.26	0.352
3-Way interactions	1	5.062	5.062	5.062	2.08	0.188
Residual error	8	19.500	19.500	2.437		
Pure error	8	19.500	19.500	2.437		
Total	15	92.938				

for fitting a model having only the three main effects. The row labeled "2-Way Interactions" refers to *AB*, *AC*, and *BC*, and the sequential sum of squares reported here is the increase in the model sum of squares if the interaction terms are added to a model containing only the main effects. Similarly, the sequential sum of squares for the three-way interaction is the increase in the model sum of squares that results from adding the term *ABC* to a model containing all other effects.

The column headed "Adj SS" (an abbreviation for adjusted sum of squares) reports how much the model sum of squares increases when each group of terms is added to a model that contains *all* the other terms. Now because any 2^k design with an equal number of replicates in each cell is an **orthogonal design**, the adjusted sum of squares equals the sequential sum of squares. Therefore, the *F*-tests for each row in the computer analysis of variance in Table 14.17 are testing the significance of each group of terms (main effects, two-factor interactions, and three-factor interactions) as if they were the last terms to be included in the model. Clearly, only the main effect terms are significant. The *t*-tests on the individual factor effects indicate that feed rate and depth of cut have large main effects, and there may be some mild interaction between these two factors. Therefore, the computer output agrees with the results given previously.

Models and Residual Analysis We may obtain the residuals from a 2^k design by using the method demonstrated earlier for the 2^2 design. As an example, consider the surface roughness experiment. The three largest effects are *A, B,* and the *AB* interaction. The regression model used to obtain the predicted values is

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon$$

where x_1 represents factor *A*, x_2 represents factor *B*, and x_1x_2 represents the *AB* interaction. The regression coefficients β_1, β_2, and β_{12} are estimated by one-half the corresponding effect estimates, and β_0 is the grand average. Thus,

$$\begin{aligned}\hat{y} &= 11.0625 + \left(\frac{3.375}{2}\right)x_1 + \left(\frac{1.625}{2}\right)x_2 + \left(\frac{1.375}{2}\right)x_1x_2 \\ &= 11.0625 + 1.6875x_1 + 0.8125x_2 + 0.6875x_1x_2\end{aligned}$$

Note that the regression coefficients are presented in the upper panel of Table 14.17. The predicted values would be obtained by substituting the low and high levels of *A* and *B* into this equation. To illustrate this, at the treatment combination where *A*, *B*, and *C* are all at the low level, the predicted value is

$$\hat{y} = 11.0625 + 1.6875(-1) + 0.8125(-1) + 0.6875(-1)(-1) = 9.25$$

Because the observed values at this run are 9 and 7, the residuals are $9 - 9.25 = -0.25$ and $7 - 9.25 = -2.25$. Residuals for the other 14 runs are obtained similarly.

See a normal probability plot of the residuals in Figure 14.19. Because the residuals lie approximately along a straight line, we do not suspect any problem with normality in the data. There are no indications of severe outliers. It would also be helpful to plot the residuals versus the predicted values and against each of the factors *A*, *B*, and *C*.

Projection of 2^k Designs Any 2^k design collapses or projects into another 2^k design in fewer variables if one or more of the original factors are dropped. Sometimes this can provide additional insight into the remaining factors. For example, consider the surface roughness experiment. Because factor *C* and all its interactions are negligible, we could eliminate factor *C* from the design. The result is to collapse the cube in Figure 14.17 into a square in the $A - B$ plane; therefore, each of the four runs in the new design has four replicates. In general, if we delete *h* factors so that $r = k - h$ factors remain, the original 2^k design with *n* replicates projects into a 2^r design with $n2^h$ replicates.

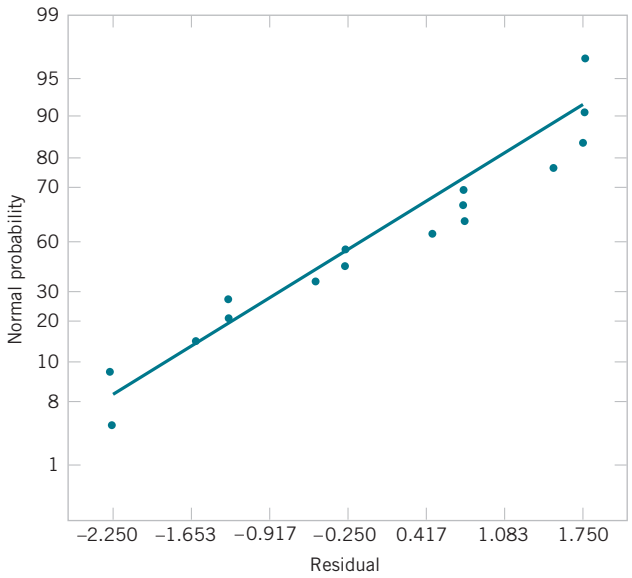


FIGURE 14.19
Normal probability plot of residuals from the surface roughness experiment.

14.6 Single Replicate of the 2^k Design

As the number of factors in a factorial experiment increases, the number of effects that can be estimated also increases. For example, a 2^4 experiment has 4 main effects, 6 two-factor interactions, 4 three-factor interactions, and 1 four-factor interaction, and a 2^6 experiment has 6 main effects, 15 two-factor interactions, 20 three-factor interactions, 15 four-factor interactions, 6 five-factor interactions, and 1 six-factor interaction. In most situations, the **sparsity of effects principle** applies; that is, the system is usually dominated by the main effects and low-order interactions. The three-factor and higher-order interactions are usually negligible. Therefore, when the number of factors is moderately large, say, $k \geq 4$ or 5, a common practice is to run only a single replicate of the 2^k design and then pool or combine the higher-order interactions as an estimate of error. Sometimes a single replicate of a 2^k design is called an **unreplicated** 2^k factorial design.

When analyzing data from unreplicated factorial designs, occasionally real high-order interactions occur. The use of an error mean square obtained by pooling high-order interactions is inappropriate in these cases. A simple method of analysis can be used to overcome this problem. Construct a plot of the estimates of the effects on a normal probability scale. The effects that are negligible are normally distributed with mean zero and variance σ^2 and tend to fall along a straight line on this plot, whereas significant effects has nonzero means and will not lie along the straight line. We illustrate this method in Example 14.5.

EXAMPLE 14.5 | Plasma Etch

An article in *Solid State Technology* ["Orthogonal Design for Process Optimization and Its Application in Plasma Etching" (May 1987, pp. 127–132)] describes the application of factorial designs in developing a nitride etch process on a single-wafer plasma etcher. The process uses C_2F_6 as the reactant gas. It is possible to vary the gas flow, the power applied to the cathode, the pressure in the reactor chamber, and the spacing between the anode and the cathode (gap). Several response variables would usually be of interest in this process, but in this example, we concentrate on etch rate for silicon nitride.

We use a single replicate of a 2^4 design to investigate this process. Because it is unlikely that the three- and four-factor interactions are significant, we tentatively plan to combine them as an estimate of error. The factor levels used in the design follow:

	Design Factor			
Level	Gap (cm)	Pressure (mTorr)	C_2F_6 Flow (SCCM)	Power (w)
Low (−)	0.80	450	125	275
High (+)	1.20	550	200	325

Refer to Table 14.18 for the data from the 16 runs of the 2^4 design. For example, the estimate of factor A is

$$\begin{aligned} A &= \frac{1}{8}[a + ab + ac + abc + ad + abd + acd + abcd \\ &\quad - (1) - b - c - bc - d - bd - cd - bcd] \\ &= \frac{1}{8}[669 + 650 + 642 + 635 + 749 + 868 + 860 + 729 \\ &\quad - 550 - 604 - 633 - 601 - 1037 - 1052 - 1075 \\ &\quad - 1063] \\ &= -101.625 \end{aligned}$$

Thus, the effect of increasing the gap between the anode and the cathode from 0.80 to 1.20 centimeters is to decrease the etch rate by 101.625 angstroms per minute.

It is easy to verify (using computer software, for example) that the complete set of effect estimates is

A	$= -101.625$	AD	$= -153.625$
B	$= -1.625$	BD	$= -0.625$
AB	$= -7.875$	ABD	$= 4.125$
C	$= 7.375$	CD	$= -2.125$
AC	$= -24.875$	ACD	$= 5.625$
BC	$= -43.875$	BCD	$= -25.375$
ABC	$= -15.625$	$ABCD$	$= -40.125$
D	$= 306.125$		

The normal probability plot of these effects from the plasma etch experiment is shown in Figure 14.20. Clearly, the main effects of A and D and the AD interaction are significant because they fall far from the line passing through the other points. The analysis, summarized in Table 14.19, confirms these findings. Notice that in the analysis of variance we have pooled the three- and four-factor interactions to form the error mean square. If the normal probability plot had indicated that any of these interactions were important, they would not have been included in the error term.

TABLE 14.18 **The 2^4 Design for the Plasma Etch Experiment**

A (Gap)	B (Pressure)	C (C_2F_6 Flow)	D (Power)	Etch Rate (Å/min)
−1	−1	−1	−1	550
1	−1	−1	−1	669
−1	1	−1	−1	604
1	1	−1	−1	650
−1	−1	1	−1	633
1	−1	1	−1	642
−1	1	1	−1	601
1	1	1	−1	635
−1	−1	−1	1	1037
1	−1	−1	1	749
−1	1	−1	1	1052
1	1	−1	1	868
−1	−1	1	1	1075
1	−1	1	1	860
−1	1	1	1	1063
1	1	1	1	729

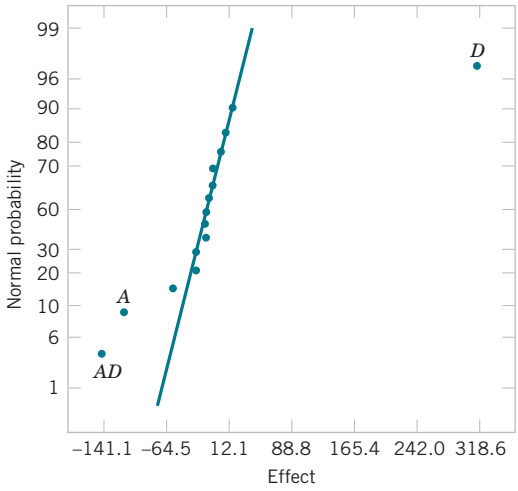


FIGURE 14.20
Normal probability plot of effects from the plasma etch experiment.

Practical Interpretation: Because $A = -101.625$, the effect of increasing the gap between the cathode and anode is to decrease the etch rate. However, $D = 306.125$; thus, applying higher power levels increase the etch rate. Figure 14.21 is a plot of the AD interaction. This plot indicates that the effect of changing the gap width at low power settings is small but that increasing the gap at high power settings dramatically reduces the etch rate. High etch rates are obtained at high power settings and narrow gap widths.

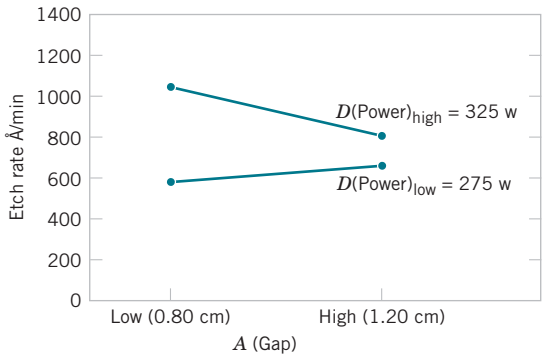


FIGURE 14.21
***AD* (gap-power) interaction from the plasma etch experiment.**

TABLE 14.19 Analysis for the Plasma Etch Experiment

Term	Effect	Coefficient	SE Coefficient	t	P-Value
Constant		776.06	11.28	68.77	0.000
A	−101.62	−50.81	11.28	−4.50	0.006
B	−1.62	−0.81	11.28	−0.07	0.945
C	7.37	3.69	11.28	0.33	0.757
D	306.12	153.06	11.28	13.56	0.000
AB	−7.88	−3.94	11.28	−0.35	0.741
AC	−24.88	−12.44	11.28	−1.10	0.321
AD	−153.62	−76.81	11.28	−6.81	0.001
BC	−43.87	−21.94	11.28	−1.94	0.109
BD	−0.63	−0.31	11.28	−0.03	0.979
CD	−2.13	−1.06	11.28	−0.09	0.929

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-Value
A	41,310.563	1	41,310.563	20.28	0.0064
B	10.563	1	10.563	<1	—
C	217.563	1	217.563	<1	—
D	374,850.063	1	374,850.063	183.99	0.0000
AB	248.063	1	248.063	<1	—
AC	2,475.063	1	2,475.063	1.21	0.3206
AD	94,402.563	1	94,402.563	46.34	0.0010
BC	7,700.063	1	7,700.063	3.78	0.1095
BD	1.563	1	1.563	<1	—
CD	18.063	1	18.063	<1	—
Error	10,186.813	5	2,037.363		
Total	531,420.938	15			

The residuals from the experiment in Example 14.5 can be obtained from the regression model

$$\hat{y} = 776.0625 - \left(\frac{101.625}{2}\right)x_1 + \left(\frac{306.125}{2}\right)x_4 - \left(\frac{153.625}{2}\right)x_1x_4$$

For example, when both A and D are at the low level, the predicted value is

$$\hat{y} = 776.0625 - \left(\frac{101.625}{2}\right)(-1) + \left(\frac{306.125}{2}\right)(-1) - \left(\frac{153.625}{2}\right)(-1)(-1) = 597$$

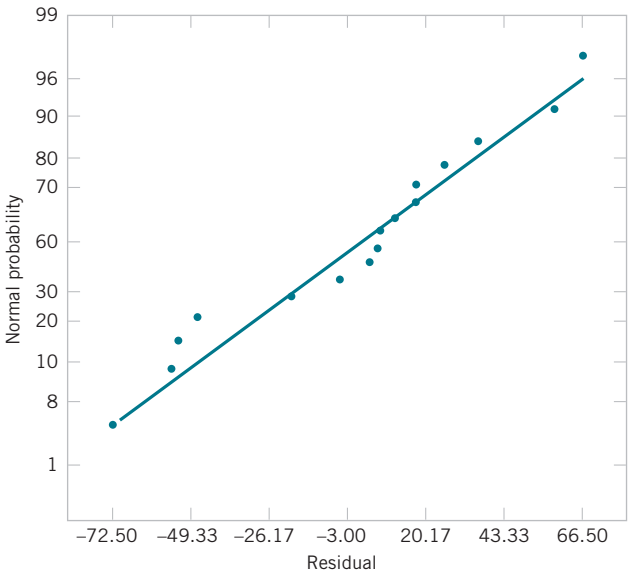


FIGURE 14.22
Normal probability plot of residuals from the plasma etch experiment.

and the four residuals at this treatment combination are

$$e_1 = 550 - 597 = -47 \qquad e_2 = 604 - 597 = 7$$
$$e_3 = 633 - 597 = 36 \qquad e_4 = 601 - 597 = 4$$

The residuals at the other three treatment combinations (A high, D low), (A low, D high), and (A high, D high) are obtained similarly. A normal probability plot of the residuals is shown in Figure 14.22. The plot is satisfactory.

14.7 Addition of Center Points to a 2^k Design

A potential concern in the use of two-level factorial designs is the assumption of linearity in the factor effects. Of course, perfect linearity is unnecessary, and the 2^k system works quite well even when the linearity assumption holds only approximately. However, a method of replicating certain points in the 2^k factorial provides protection against curvature and allows an independent estimate of error to be obtained. The method consists of adding **center points** to the 2^k design. These consist of n_C replicates run at the point $x_i = 0$ ($i = 1, 2, \ldots, k$). One important reason for adding the replicate runs at the design center is that center points do not affect the usual effects estimates in a 2^k design. We assume that the k factors are quantitative.

To illustrate the approach, consider a 2^2 design with one observation at each of the factorial points $(-, -)$, $(+, -)$, $(-, +)$, and $(+, +)$ and n_C observations at the center points $(0, 0)$. Figure 14.23 illustrates the situation. Let $\bar{y}_F$ be the average of the four runs at the four factorial points, and let $\bar{y}_C$ be the average of the n_C run at the center point. If the difference $\bar{y}_F - \bar{y}_C$ is small, the center points lie on or near the plane passing through the factorial points, and there is no curvature. On the other hand, if $\bar{y}_F - \bar{y}_C$ is large, curvature is present.

Similar to factorial effects, a test for curvature can be based on a F-statistic or an equivalent t-statistic. A single degree-of-freedom sum of squares for curvature is compared to MS_E to produce the F-statistic. Alternatively, a t-statistic similar to the one used to compare two means can be computed. A coefficient for curvature is defined to be $\bar{y}_F - \bar{y}_C$, and σ is estimated by the square root of MS_E.

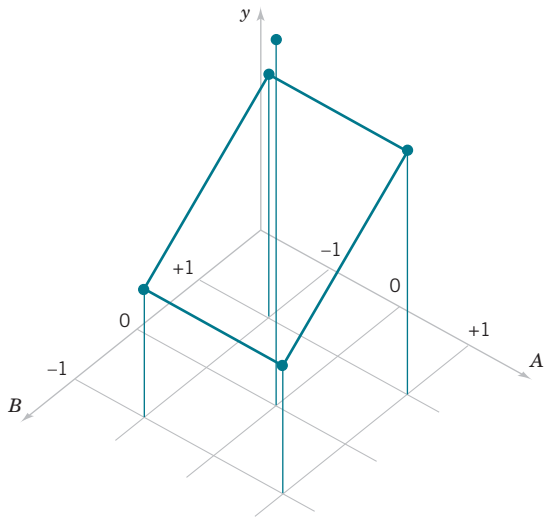


FIGURE 14.23

A 2^2 design with center points.

Curvature Sum of Squares and *t*-Statistic

The sum of squares for curvature is

$$SS_{\text{Curvature}} = \frac{n_F n_C (\bar{y}_F - \bar{y}_C)^2}{n_F + n_C} = \left(\frac{\bar{y}_F - \bar{y}_C}{\sqrt{\frac{1}{n_F} + \frac{1}{n_C}}} \right)^2$$

and the t-statistic to test for curvature is

$$t = \frac{\bar{y}_F - \bar{y}_C}{\hat{\sigma}\sqrt{\frac{1}{n_F} + \frac{1}{n_C}}} \tag{14.14}$$

where, in general, n_F is the number of factorial design points. The $SS_{\text{Curvature}}$ may be compared to the error mean square to produce the F-test for curvature. Notice that, similar to the test for other effects, the square of the t-statistic equals the F-statistic.

When points are added to the center of the 2^k design, the model we may entertain is

$$Y = \beta_0 + \sum_{j=1}^{k} \beta_0 x_j + \sum_{i<j} \sum \beta_{ij} x_i x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \epsilon$$

where the β_{jj} are pure quadratic effects. The test for curvature actually tests the hypotheses

$$H_0: \sum_{j=1}^{k} \beta_{jj} = 0 \qquad H_1: \sum_{j=1}^{k} \beta_{jj} \neq 0$$

Furthermore, if the factorial points in the design are unreplicated, we may use the n_C center points to construct an estimate of error with $n_C - 1$ degrees of freedom.

EXAMPLE 14.6 | Process Yield

A chemical engineer is studying the percentage of conversion or yield of a process. There are two variables of interest, reaction time and reaction temperature. Because she is uncertain about the assumption of linearity over the region of exploration, the engineer decides to conduct a 2^2 design (with a single replicate of each factorial run) augmented with five center points. The design and the yield data are shown in Figure 14.24.

Table 14.20 summarizes the analysis for this experiment. The mean square error is calculated from the center points as follows:

$$MS_E = \frac{SS_E}{n_C - 1} = \frac{\sum_{\text{Center points}} (y_i - \bar{y}_C)^2}{n_C - 1} = \frac{\sum_{i=1}^{5} (y_i - 40.46)^2}{4}$$
$$= \frac{0.1720}{4} = 0.0430$$

The average of the points in the factorial portion of the design is $\bar{y}_F = 40.425$, and the average of the points at the center is $\bar{y}_C = 40.46$. The difference $\bar{y}_F - \bar{y}_C = 40.425 - 40.46 = -0.035$ appears to be small. The curvature sum of squares in the analysis of variance table is computed from Equation 14.14 as follows:

$$SS_{\text{Curvature}} = \frac{n_F n_C (\bar{y}_F - \bar{y}_C)^2}{n_F + n_C} = \frac{(4)(5)(-0.035)^2}{4 + 5} = 0.0027$$

The coefficient for curvature is $\bar{y}_F - \bar{y}_C = -0.035$, and the t-statistic to test for curvature is

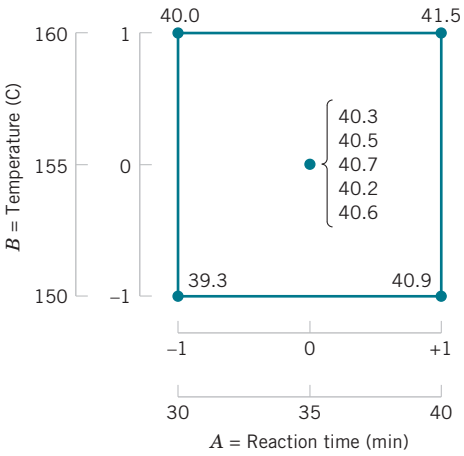


FIGURE 14.24

The 2^2 design with five center points for the process yield experiment.

$$t = \frac{-0.035}{\sqrt{0.043\left(\frac{1}{4} + \frac{1}{5}\right)}} = -0.25$$

Practical Interpretation: The analysis of variance indicates that both factors exhibit significant main effects, that there is no interaction, and that there is no evidence of curvature in the response over the region of exploration. That is, the null hypothesis H_0: $\sum_{j=1}^{k} \beta_{jj} = 0$ cannot be rejected.

TABLE 14.20 **Analysis for the Process Yield Experiment with Center Points**

Term	Effect	Coefficient	SE Coefficient	t	P-Value
Constant		40.4250	0.1037	389.89	0.000
A	1.5500	0.7750	0.1037	7.47	0.002
B	0.6500	0.3250	0.1037	3.13	0.035
AB	−0.0500	−0.0250	0.1037	−0.24	0.821
Ct Pt		−0.0350	0.1391	−0.25	0.814

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-Value
A (Time)	2.4025	1	2.4025	55.87	0.0017
B (Temperature)	0.4225	1	0.4225	9.83	0.0350
AB	0.0025	1	0.0025	0.06	0.8237
Curvature	0.0027	1	0.0027	0.06	0.8163
Error	0.1720	4	0.0430		
Total	3.0022	8			

14.8 Blocking and Confounding in the 2^k Design

Running all the observations in a 2^k factorial design under homogeneous conditions is often impossible. **Blocking** is the appropriate design technique for this general situation. However, in many situations, the block size is smaller than the number of runs in the complete replicate. In these cases, **confounding** is a useful procedure for running the 2^k design in 2^p blocks where the number of runs in a block is less than the number of treatment combinations in one complete replicate. The technique causes certain interaction effects to be indistinguishable from blocks or **confounded with blocks**. We illustrate confounding in the 2^k factorial design in 2^p blocks where $p < k$.

Consider a 2^2 design. Suppose that each of the $2^2 = 4$ treatment combinations requires 4 hours of laboratory analysis. Thus, 2 days are required to perform the experiment. If days are considered as blocks, we must assign two of the four treatment combinations to each day.

See this design in Figure 14.25. Notice that block 1 contains the treatment combinations (1) and ab and that block 2 contains a and b. The contrasts for estimating the main effects of factors A and B are

$$\text{Contrast}_A = ab + a - b - (1) \qquad \text{Contrast}_B = ab + b - a - (1)$$

Note that these contrasts are unaffected by blocking because in each contrast there is one plus and one minus treatment combination from each block. That is, any difference between block 1 and block 2 that increases the readings in one block by an additive constant cancels out. The contrast for the AB interaction is

$$\text{Contrast}_{AB} = ab + (1)b - a - b$$

Because the two treatment combinations with the plus signs, ab and (1), are in block 1 and the two with the minus signs, a and b, are in block 2, the block effect and the AB interaction are identical. That is, the AB interaction is confounded with blocks. The reason for this is apparent from the table of plus and minus signs for the 2^2 design in Table 14.11. From the table, we see that all treatment combinations that have a plus on AB are assigned to block 1 whereas all treatment combinations that have a minus sign on AB are assigned to block 2.

This scheme can be used to confound any 2^k design in two blocks. As a second example, consider a 2^3 design, run in two blocks. From the table of plus and minus signs in Table 14.14, we assign the treatment combinations that are minus in the ABC column to block 1 and those that are plus in the ABC column to block 2. The resulting design is shown in Figure 14.26.

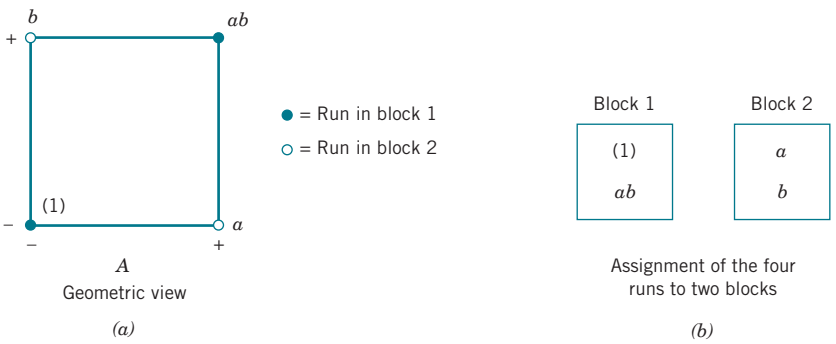


FIGURE 14.25

A 2^2 design in two blocks. (a) Geometric view. (b) Assignment of the four runs to two blocks.

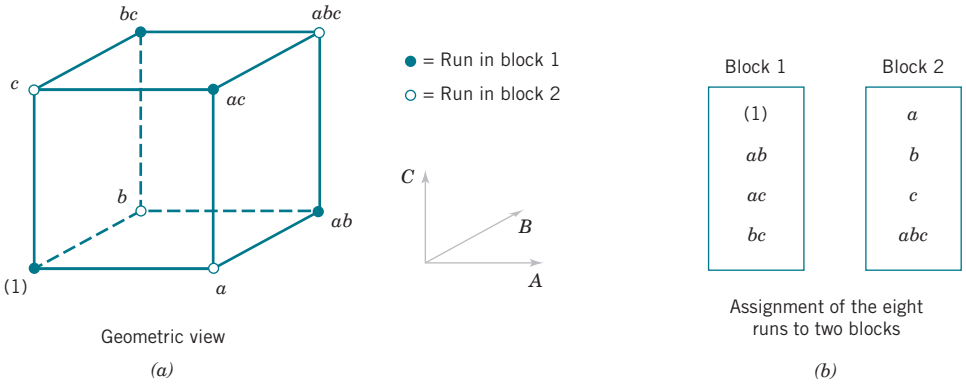


FIGURE 14.26

The 2^3 design in two blocks with ABC confounded. (a) Geometric view. (b) Assignment of the eight runs to two blocks.

There is a more general method of constructing the blocks. The method employs a **defining contrast**, say,

$$L = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n \tag{14.15}$$

where x_i is the level of the ith factor appearing in a treatment combination and α_i is the exponent appearing on the ith factor in the effect that is to be confounded with blocks. For the 2^k system, we have either $\alpha_i = 0$ or 1 and either $x_i = 0$ (low level) or $x_i = 1$ (high level). Treatment combinations that produce the same value of L (modulus 2) are placed in the same block. Because the only possible values of L (mod 2) are 0 and 1, this assigns the 2^k treatment combinations to exactly two blocks.

As an example, consider the 2^3 design with ABC confounded with blocks. Here x_1 corresponds to A, x_2 to B, x_3 to C, and $\alpha_1 = \alpha_2 = \alpha_3 = 1$. Thus, the defining contrast that would be used to confound ABC with blocks is

$$L = x_1 + x_2 + x_3$$

To assign the treatment combinations to the two blocks, we substitute the treatment combinations into the defining contrast as follows:

$$\begin{aligned}
(1)&: \ L = 1(0) + 1(0) + 1(0) = 0 = 0 \pmod 2 & c&: \ L = 1(0) + 1(0) + 1(1) = 1 = 1 \pmod 2\\
a&: \ L = 1(1) + 1(0) + 1(0) = 1 = 1 \pmod 2 & ac&: \ L = 1(1) + 1(0) + 1(1) = 2 = 0 \pmod 2\\
b&: \ L = 1(0) + 1(1) + 1(0) = 1 = 1 \pmod 2 & bc&: \ L = 1(0) + 1(1) + 1(1) = 2 = 0 \pmod 2\\
ab&: \ L = 1(1) + 1(1) + 1(0) = 2 = 0 \pmod 2 & abc&: \ L = 1(1) + 1(1) + 1(1) = 3 = 1 \pmod 2
\end{aligned}$$

Thus (1), ab, ac, and bc are run in block 1, and a, b, c, and abc are run in block 2. This same design is shown in Figure 14.26.

A shortcut method is useful in constructing these designs. The block containing the treatment combination (1) is called the **principal block**. Any element [except (1)] in the principal block may be generated by multiplying two other elements in the principal block modulus 2 on the exponents. For example, consider the principal block of the 2^3 design with ABC confounded, shown in Figure 14.26. Note that

$$\begin{aligned}
ab \cdot ac &= a^2bc = bc\\
ab \cdot bc &= ab^2c = ac\\
ac \cdot bc &= abc^2 = ab
\end{aligned}$$

Treatment combinations in the other block (or blocks) may be generated by multiplying one element in the new block by each element in the principal block modulus 2 on the exponents. For the 2^3 with ABC confounded, because the principal block is (1), ab, ac, and bc, we know that the treatment combination b is in the other block. Thus, elements of this second block are

$$\begin{aligned} b\cdot(1) &= b \\ b\cdot ab = ab^2 &= a \\ b\cdot ac &= abc \\ b\cdot bc = b^2c &= c \end{aligned}$$

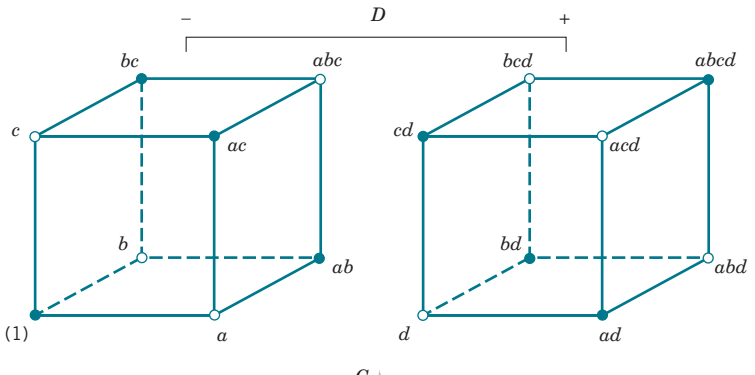
EXAMPLE 14.7 | Missile Miss Distance

An experiment is performed to investigate the effect of four factors on the terminal miss distance of a shoulder-fired ground-to-air missile. The four factors are target type (A), seeker type (B), target altitude (C), and target range (D). Each factor may be conveniently run at two levels, and the optical tracking system allows terminal miss distance to be measured to the nearest foot. Two different operators or gunners are used in the flight test and, because there may be differences between operators, the test engineers decided to conduct the 2^4 design in two blocks with $ABCD$ confounded. Thus, the defining contrast is

$$L = x_1 + x_2 + x_3 + x_4$$

The experimental design and the resulting data are shown in Figure 14.27. The effect estimates obtained from computer software are shown in Table 14.21. A normal probability plot of the effects in Figure 14.28 reveals that A (target type), D (target range), AD, and AC have large effects. A confirming analysis, pooling the three-factor interactions as error, is shown in Table 14.22.

Practical Interpretation: Because the AC and AD interactions are significant, it is logical to conclude that A (target type), C (target altitude), and D (target range) all have important effects on the miss distance and that there are interactions between target type and altitude and target type and range. Notice that the $ABCD$ effect is treated as blocks in this analysis.



Geometric view

(a)

Block 1		Block 2	
(1)	= 3	a	= 7
ab	= 7	b	= 5
ac	= 6	c	= 6
bc	= 8	d	= 4
ad	= 10	abc	= 6
bd	= 4	bcd	= 7
cd	= 8	acd	= 9
$abcd$	= 9	abd	= 12

Assignment of the sixteen runs to two blocks

(b)

FIGURE 14.27

The 2^4 design in two blocks for Example 14.7. (a) Geometric view. (b) Assignment of the 16 runs to two blocks.

TABLE 14.21 **Effect Estimates from Computer Output for the Missile Miss Distance Experiment**

Estimated Effects and Coefficients for Distance		
Term	Effect	Coefficient
Constant		6.938
Block (*ABCD*)		0.063
A	2.625	1.312
B	0.625	0.313
C	0.875	0.438
D	1.875	0.938
AB	−0.125	−0.063
AC	−2.375	−1.187
AD	1.625	0.813
BC	−0.375	−0.188
BD	−0.375	−0.187
CD	−0.125	−0.062
ABC	−0.125	−0.063
ABD	0.875	0.438
ACD	−0.375	−0.187
BCD	−0.375	−0.187

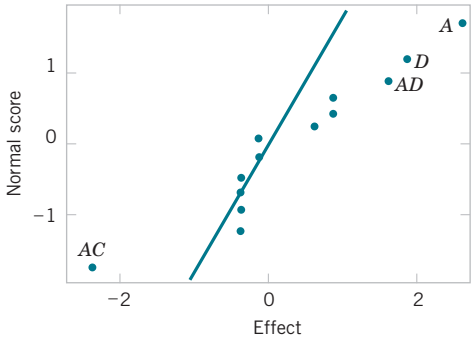


FIGURE 14.28
Normal probability plot of the effects for the Missile Miss Distance Experiment.

TABLE 14.22 **Analysis for the Missile Miss Distance Experiment**

Term	Effect	Coefficient	SE Coefficient	*t*	*P*-Value
Constant		6.938	0.2577	26.92	0.000
Blocks (*ABCD*)		0.063	0.2577	0.24	0.820
A	2.625	1.312	0.2577	5.09	0.007
B	0.625	0.313	0.2577	1.21	0.292
C	0.875	0.437	0.2577	1.70	0.165
D	1.875	0.938	0.2577	3.64	0.022
AB	−0.125	−0.062	0.2577	−0.24	0.820
AC	−2.375	−1.188	0.2577	−4.61	0.010
AD	1.625	0.813	0.2577	3.15	0.034
BC	−0.375	−0.187	0.2577	−0.73	0.507
BD	−0.375	−0.188	0.2577	−0.73	0.507
CD	−0.125	−0.063	0.2577	−0.24	0.820

TABLE 14.22 **Analysis for the Missile Miss Distance Experiment *(Continued)***

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	*P*-Value
Blocks (*ABCD*)	0.0625	1	0.0625	0.06	—
A	27.5625	1	27.5625	25.94	0.0070
B	1.5625	1	1.5625	1.47	0.2920
C	3.0625	1	3.0625	2.88	0.1648
D	14.0625	1	14.0625	13.24	0.0220
AB	0.0625	1	0.0625	0.06	—
AC	22.5625	1	22.5625	21.24	0.0100
AD	10.5625	1	10.5625	9.94	0.0344
BC	0.5625	1	0.5625	0.53	—
BD	0.5625	1	0.5625	0.53	—
CD	0.0625	1	0.0625	0.06	—
Error (*ABC* + *ABD* + *ACD* + *BCD*)	4.2500	4	1.0625		
Total	84.9375	15			

It is possible to confound the 2^k design in four blocks of 2^{k-2} observations each. To construct the design, two effects are chosen to confound with blocks, and their defining contrasts are obtained. A third effect, the **generalized interaction** of the two effects initially chosen, is also confounded with blocks. The generalized interaction of two effects is found by multiplying their respective letters and reducing the exponents modulus 2.

For example, consider the 2^4 design in four blocks. If *AC* and *BD* are confounded with blocks, their generalized interaction is $(AC)(BD) = ABCD$. The design is constructed by using the defining contrasts for *AC* and *BD*:

$$L_1 = x_1 + x_3 \qquad L_2 = x_2 + x_4$$

It is easy to verify that the four blocks are

Block 1 $L_1 = 0, L_2 = 0$	Block 2 $L_1 = 1, L_2 = 0$	Block 3 $L_1 = 0, L_2 = 1$	Block 4 $L_1 = 1, L_2 = 1$
(1)	*a*	*b*	*ab*
ac	*c*	*abc*	*bc*
bd	*abd*	*d*	*ad*
abcd	*bcd*	*acd*	*cd*

This general procedure can be extended to confounding the 2^k design in 2^p blocks where $p < k$. Start by selecting p effects to be confounded such that no effect chosen is a generalized interaction of the others. Then the blocks can be constructed from the p defining contrasts $L_1, L_2, \ldots, L_p$ that are associated with these effects. In addition to the p effects chosen to be confounded, exactly $2^p - p - 1$ additional effects are confounded with blocks; these are the generalized interactions of the original p effects chosen. Care should be taken so as not to confound effects of potential interest.

For more information on confounding in the 2^k factorial design, refer to Montgomery (2017) for guidelines on selecting factors to confound with blocks so that main effects and low-order interactions are not confounded. In particular, that book has a table of suggested confounding schemes for designs with up to seven factors and a range of block sizes, some of which are as small as two runs.

14.9 One-Half Fraction of the 2^k Design

As the number of factors in a 2^k factorial design increases, the number of runs required increases rapidly. For example, a 2^5 requires 32 runs. In this design, only 5 degrees of freedom correspond to main effects, and 10 degrees of freedom correspond to two-factor interactions. Sixteen of the 31 degrees of freedom are used to estimate high-order interactions—that is, three-factor and higher-order interactions. Often there is little interest in these high-order interactions, particularly when we first begin to study a process or system. If we can assume that certain high-order interactions are negligible, a **fractional factorial design** involving fewer than the complete set of 2^k runs can be used to obtain information on the main effects and low-order interactions. In this section, we introduce fractional replications of the 2^k design.

A major use of fractional factorials is in *screening experiments*. These are experiments in which many factors are considered with the purpose of identifying those factors (if any) that have large effects. Screening experiments are usually performed in the early stages of a project when it is likely that many of the factors initially considered have little or no effect on the response. The factors that are identified as important are then investigated more thoroughly in subsequent experiments.

A one-half fraction of the 2^k design contains 2^{k-1} runs and is often called a 2^{k-1} fractional factorial design. As an example, consider the 2^{3-1} design—that is, a one-half fraction of the 2^3. This design has only four runs, in contrast to the full factorial that would require eight runs. The table of plus and minus signs for the 2^3 design is shown in Table 14.23. Suppose that we select the four treatment combinations a, b, c, and abc as our one-half fraction. These treatment combinations are shown in the top half of Table 14.23 and in Figure 14.29(a).

Notice that the 2^{3-1} design is formed by selecting only those treatment combinations that yield a plus on the ABC effect. Thus, ABC is called the **generator** of this particular fraction. Furthermore, the identity element I is also a plus for the four runs, so we call

$$I = ABC$$

the **defining relation** for the design.

TABLE 14.23 Plus and Minus Signs for the 2^3 Factorial Design

Treatment Combination	Factorial Effect							
	I	A	B	C	AB	AC	BC	ABC
a	+	+	−	−	−	−	+	+
b	+	−	+	−	−	+	−	+
c	+	−	−	+	+	−	−	+
abc	+	+	+	+	+	+	+	+
ab	+	+	+	−	+	−	−	−
ac	+	+	−	+	−	+	−	−
bc	+	−	+	+	−	−	+	−
(1)	+	−	−	−	+	+	+	−

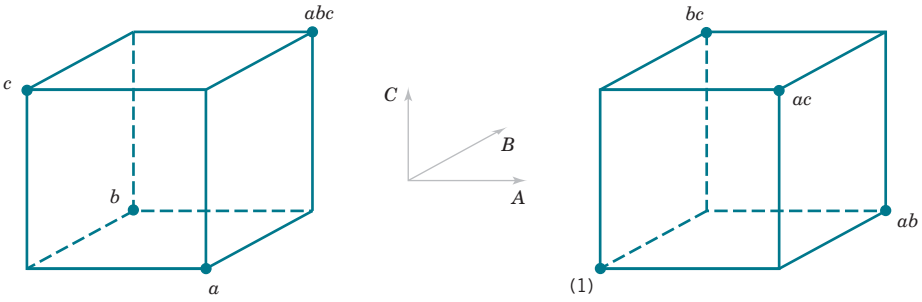


FIGURE 14.29

The one-half fractions of the 2^3 design. (a) The principal fraction, $I = +ABC$. (b) The alternate fraction, $I = -ABC$.

The treatment combinations in the 2^{3-1} design yields three degrees of freedom associated with the main effects. From the upper half of Table 14.23, we obtain the estimates of the main effects as linear combinations of the observations, say,

$$A = \tfrac{1}{2}[a - b - c + abc] \qquad B = \tfrac{1}{2}[-a + b - c + abc] \qquad C = \tfrac{1}{2}[-a - b + c + abc]$$

It is also easy to verify that the estimates of the two-factor interactions should be the following linear combinations of the observations:

$$BC = \tfrac{1}{2}[a - b - c + abc] \qquad AC = \tfrac{1}{2}[-a + b - c + abc] \qquad AB = \tfrac{1}{2}[-a - b + c + abc]$$

Thus, the linear combination of observations in column A, called ℓ_A, estimates both the main effect of A and the BC interaction. That is, the linear combination ℓ_A estimates the sum of these two effects $A + BC$. Similarly, ℓ_B estimates $B + AC$, and ℓ_C estimates $C + AB$. Two or more effects that have this property are called **aliases**. In our 2^{3-1} design, A and BC are aliases, B and AC are aliases, and C and AB are aliases. Aliasing is the direct result of fractional replication. In many practical situations, it is possible to select the fraction so that the main effects and low-order interactions that are of interest are aliased only with high-order interactions (which are probably negligible).

The alias structure for this design is found by using the defining relation $I = ABC$. Multiplying any effect by the defining relation yields the aliases for that effect. In our example, the alias of A is

$$A = A \cdot ABC = A^2BC = BC$$

because $A \cdot I = A$ and $A^2 = I$. The aliases of B and C are

$$B = B \cdot ABC = AB^2C = AC \qquad \text{and} \qquad C = C \cdot ABC = ABC^2 = AB$$

Now suppose that we had chosen the other one-half fraction, that is, the treatment combinations in Table 14.23 associated with minus on ABC. See these four runs in the lower half of Table 14.23 and in Figure 14.29(b). The defining relation for this design is $I = -ABC$. The aliases are $A = -BC$, $B = -AC$, and $C = -AB$. Thus, estimates of A, B, and C that result from this fraction really estimate $A - BC$, $B - AC$, and $C - AB$. In practice, it usually does not matter which one-half fraction we select. The fraction with the plus sign in the defining relation is usually called the **principal fraction**, and the other fraction is usually called the **alternate fraction**.

Note that if we had chosen AB as the generator for the fractional factorial $A = A \cdot AB = B$ and the two main effects of A and B would be aliased. This typically loses important information.

Sometimes we use **sequences** of fractional factorial designs to estimate effects. For example, suppose that we had run the principal fraction of the 2^{3-1} design with generator ABC. From this design, we have the following effect estimates:

$$\ell_A = A + BC \qquad \ell_B = B + AC \qquad \ell_C = C + AB$$

Suppose that we are willing at this point to assume that the two-factor interactions are negligible. If they are, the 2^{3-1} design has produced estimates of the three main effects A, B, and C. However, if after running the principal fraction, we are uncertain about the interactions, it is possible to estimate them by running the alternate fraction. The alternate fraction produces the following effect estimates:

$$\ell'_A = A - BC \qquad \ell'_B = B - AC \qquad \ell'_C = C - AB$$

We may now obtain de-aliased estimates of the main effects and two-factor interactions by adding and subtracting the linear combinations of effects estimated in the two individual fractions. For example, suppose that we want to de-alias A from the two-factor interaction BC. Because $\ell_A = A + BC$ and $\ell'_A = A - BC$, we can combine these effect estimates as follows:

$$\frac{1}{2}(\ell_A - \ell'_A) = \frac{1}{2}(A + BC + A - BC) = A$$

and

$$\frac{1}{2}(\ell_A - \ell'_A) = \frac{1}{2}(A + BC - A + BC) = BC$$

For all three pairs of effect estimates, we would obtain the following results:

Effect, i	**from $\frac{1}{2}(l_i + l'_i)$**	**from $\frac{1}{2}(l_i - l'_i)$**
$i = A$	$\frac{1}{2}(A + BC + A - BC) = A$	$\frac{1}{2}[A + BC - (A - BC)] = BC$
$i = B$	$\frac{1}{2}(B + AC + B - AC) = B$	$\frac{1}{2}[B + AC - (B - AC)] = AC$
$i = C$	$\frac{1}{2}(C + AB + C - AB) = C$	$\frac{1}{2}[C + AB - (C - AB)] = AB$

Thus, by combining a sequence of two fractional factorial designs, we can isolate both the main effects and the two-factor interactions. This property makes the fractional factorial design highly useful in experimental problems because we can run sequences of small, efficient experiments, combine information across *several* experiments, and take advantage of learning about the process we are experimenting with as we go along. This is an illustration of the concept of sequential experimentation.

A 2^{k-1} design may be constructed by writing down the treatment combinations for a full factorial with $k - 1$ factors, called the *basic design*, and then adding the kth factor by identifying its plus and minus levels with the plus and minus signs of the highest-order interaction. Therefore, a 2^{3-1} fractional factorial is constructed by writing down the basic design as a full 2^2 factorial and then equating factor C with the $\pm AB$ interaction. Thus, to construct the principal fraction, we would use $C = +AB$ as follows:

Basic Design		**Fractional Design**		
Full 2^2		**2^{3-1}, $I = +ABC$**		
A	***B***	***A***	***B***	***C = AB***
−	−	−	−	+
+	−	+	−	−
−	+	−	+	−
+	+	+	+	+

To obtain the alternate fraction, we would equate the last column to $C = -AB$.

EXAMPLE 14.8 | Plasma Etch

To illustrate the use of a one-half fraction, consider the plasma etch experiment described in Example 14.5. Suppose that we decide to use a 2^{4-1} design with $I = ABCD$ to investigate the four factors gap (A), pressure (B), C_2F_6 flow rate (C), and power setting (D). This design would be constructed by writing as the basic design a 2^3 in the factors A, B, and C and then setting the levels of the fourth factor $D = ABC$. The design and the resulting etch rates are shown in Table 14.24. The design is shown graphically in Figure 14.30.

In this design, the main effects are aliased with the three-factor interactions; note that the alias of A is

$$A \cdot I = A \cdot ABCD \qquad \text{or} \qquad A = A^2BCD = BCD$$

and similarly, $B = ACD$, $C = ABD$, and $D = ABC$. The two-factor interactions are aliased with each other. For example, the alias of AB is CD:

$$AB \cdot I = A \cdot ABCD \qquad \text{or} \qquad AB = A^2B^2CD = CD$$

The other aliases are $AC = BD$ and $AD = BC$.

The estimates of the main effects and their aliases are found using the four columns of signs in Table 14.24. For example, from column A, we obtain the estimated effect

$$\ell_A = A + BCD = \frac{1}{4}(-550 + 749 - 1052 + 650 - 1075 + 642 - 601 + 729) = -127.0$$

The other columns produce

$$\ell_B = B + ACD = 4.0 \qquad \ell_C = C + ABD = 11.5$$

and

$$\ell_D = D + ABC = 290.5$$

Clearly, ℓ_A and ℓ_D are large, and if we believe that the three-factor interactions are negligible, the main effects A (gap) and D (power setting) significantly affect etch rate.

TABLE 14.24 The 2^{4-1} Design with Defining Relation $I = ABCD$

A	B	C	$D = ABC$	Treatment Combination	Etch Rate
−	−	−	−	(1)	550
+	−	−	+	ad	749
−	+	−	+	bd	1052
+	+	−	−	ab	650
−	−	+	+	cd	1075
+	−	+	−	ac	642
−	+	+	−	bc	601
+	+	+	+	$abcd$	729

The interactions are estimated by forming the AB, AC, and AD columns and adding them to the table. For example, the signs in the AB column are +, −, −, +, +, −, −, +, and this column produces the estimate

$$\ell_{AB} = AB + CD = \frac{1}{4}(550 - 749 - 1052 + 650 + 1075 - 642 - 601 + 729) = -10$$

From the AC and AD columns we find

$$\ell_{AC} = AC + BD = -25.50 \qquad \ell_{AD} = AD + BC = -197.50$$

The ℓ_{AD} estimate is large; the most straightforward interpretation of the results is that because A and D are large, this is the AD interaction. Thus, the results obtained from the 2^{4-1} design agree with the full factorial results in Example 14.5.

Practical Interpretation: Often a fraction of a 2^k design is satisfactory when an experiment uses four or more factors.

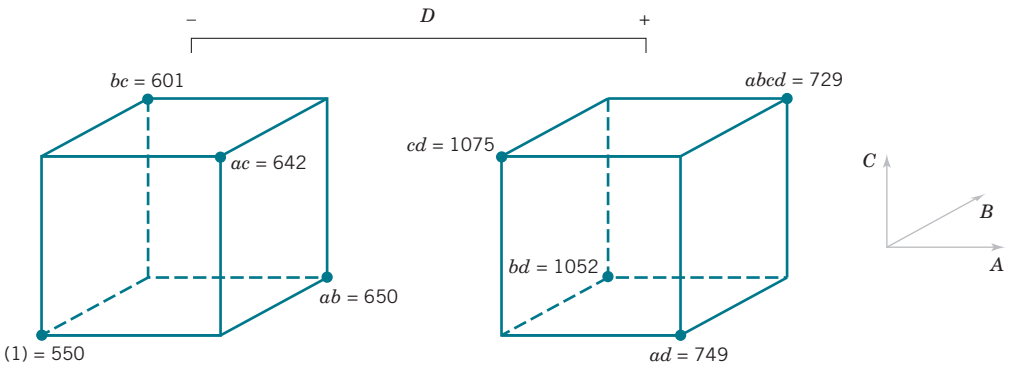


FIGURE 14.30

The 2^{4-1} design for the plasma etch experiment.

Normal Probability Plot of Effects The normal probability plot is very useful in assessing the significance of effects from a fractional factorial design, particularly when many effects are to be estimated. We strongly recommend examining this plot. Figure 14.31 presents the normal probability plot of the effects from computer software for the plasma etch experiment in Example 14.8. Notice that the A, D, and AD interaction effects stand out clearly in this graph.

Residual Analysis The residuals can be obtained from a fractional factorial by the regression model method shown previously. The residuals should be graphically analyzed as we have discussed before, both to assess the validity of the underlying model assumptions and to gain additional insight into the experimental situation.

Center Points Center points can be added to fractional factorial designs for the same reasons as before; that is, to assess curvature and provide an error estimate. The curvature analysis compares $\bar{y}_F - \bar{y}_C$ to its standard error with the same calculation as for full factorial experiments. The error estimate can be based on the replicated center points and supplemented with sums of squares from effects determined to be unimportant from the normal probability plot.

Projection of the 2^{k-1} Design If one or more factors from a one-half fraction of a 2^k can be dropped, the design projects into a full factorial design. For example, Figure 14.32 presents a 2^{3-1} design. Notice that this design projects into a full factorial in any two of the three original factors. Thus, if we think that at most two of the three factors are important, the 2^{3-1} design is an excellent design for identifying the significant factors. This **projection property** is highly useful in factor screening because it allows negligible factors to be eliminated, resulting in a stronger experiment in the active factors that remain.

In the 2^{4-1} design used in the plasma etch experiment in Example 14.8, we found that two of the four factors (B and C) could be dropped. If we eliminate these two factors, the remaining columns in Table 14.24 form a 2^2 design in the factors A and D with two replicates. This design is shown in Figure 14.33. The main effects of A and D and the strong two-factor AD interaction are clearly evident from this graph.

Design Resolution The concept of design **resolution** is a useful way to catalog fractional factorial designs according to the alias patterns they produce. Designs of resolution III, IV, and V are particularly important. The definitions of these terms and an example of each follow.

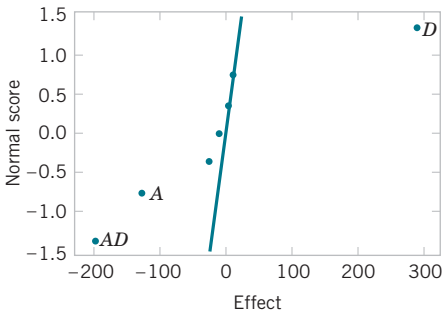


FIGURE 14.31
Normal probability plot of the effects for the plasma etch experiment in Example 14.8.

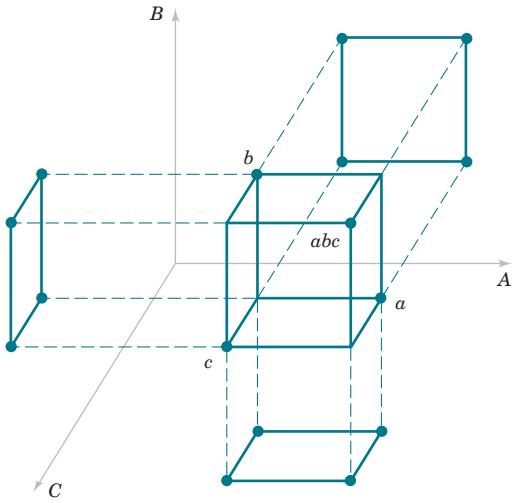


FIGURE 14.32

Projection of a 2^{3-1} design into three 2^2 designs.

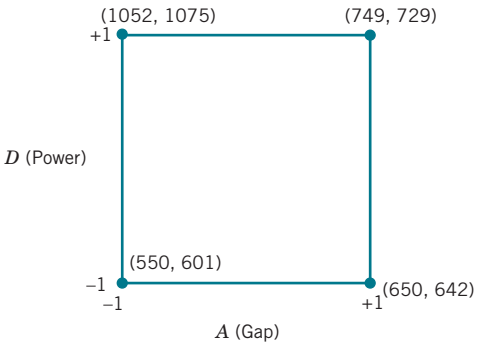


FIGURE 14.33

The 2^2 design obtained by dropping factors *B* and *C* from the plasma etch experiment in Example 14.8.

1. *Resolution III Designs.* These are designs in which no main effects are aliased with any other main effect, but main effects are aliased with two-factor interactions, and some two-factor interactions may be aliased with each other. The 2^{3-1} design with $I = ABC$ is a resolution III design. We usually employ a Roman numeral subscript to indicate design resolution; thus, this one-half fraction is a 2^{3-1}_{III} design.
2. *Resolution IV Designs.* These are designs in which no main effect is aliased with any other main effect or two-factor interactions, but two-factor interactions are aliased with each other. The 2^{4-1} design with $I = ABCD$ used in Example 14.8 is a resolution IV design (2^{4-1}_{III}).
3. *Resolution V Designs.* These are designs in which no main effect or two-factor interaction is aliased with any other main effect or two-factor interaction, but two-factor interactions are aliased with three-factor interactions. The 2^{5-1} design with $I = ABCDE$ is a resolution V design (2^{5-1}_{III}).

Resolution III and IV designs are particularly useful in factor screening experiments. A resolution IV design provides good information about main effects and some information about all two-factor interactions.

14.10 Smaller Fractions: The 2^{k-p} Fractional Factorial

Although the 2^{k-1} design is valuable in reducing the number of runs required for an experiment, we frequently find that smaller fractions provide almost as much useful information at even greater economy. In general, a 2^k design may be run in a $1/2^p$ fraction called a 2^{k-p} fractional factorial design. Thus, a 1/4 fraction is called a 2^{k-2} design, a 1/8 fraction is called a 2^{k-3} design, a 1/16 fraction a 2^{k-4} design, and so on.

To illustrate the 1/4 fraction, consider an experiment with six factors and suppose that the engineer is primarily interested in main effects but would also like to get some information about the two-factor interactions. A 2^{6-1} design would require 32 runs and would have 31 degrees of

TABLE 14.25 Alias Structure for the 2^{6-2}_{IV} Design with $I = ABCE = BCDF = ADEF$

$A = BCE = DEF = ABCDF$	$AB = CE = ACDF = BDEF$
$B = ACE = CDF = ABDEF$	$AC = BE = ABDF = CDEF$
$C = ABE = BDF = ACDEF$	$AD = EF = BCDE = ABCF$
$D = BCF = AEF = ABCDE$	$AE = BC = DF = ABCDEF$
$E = ABC = ADF = BCDEF$	$AF = DE = BCEF = ABCD$
$F = BCD = ADE = ABCEF$	$BD = CF = ACDE = ABEF$
$ABD = CDE = ACF = BEF$	$BF = CD = ACEF = ABDE$
$ACD = BDE = ABF = CEF$	

freedom for estimating effects. Because there are only six main effects and 15 two-factor interactions, the one-half fraction is inefficient—it requires too many runs. Suppose that we consider a 1/4 fraction, or a 2^{6-2} design. This design contains 16 runs and, with 15 degrees of freedom, allows all six main effects to be estimated with some capability for examining the two-factor interactions.

To generate this design, we would write down a 2^4 design in the factors A, B, C, and D as the basic design and then add two columns, for E and F. To find the new columns, we could select the two **design generators** $I = ABCE$ and $I = BCDF$. Thus, column E would be found from $E = ABC$, and column F would be $F = BCD$. That is, columns $ABCE$ and $BCDF$ are equal to the identity column. However, we know that the product of any two columns in the table of plus and minus signs for a 2^k design is just another column in the table; therefore, the product of $ABCE$ and $BCDF$ or $ABCE(BCDF) = AB^2C^2DEF = ADEF$ is also an identity column. Consequently, the **complete defining relation** for the 2^{6-2} design is

$$I = ABCE = BCDF = ADEF$$

We refer to each term in a defining relation (such as $ABCE$ above) as a *word*. To find the alias of any effect, simply multiply the effect by each word in the foregoing defining relation. For example, the alias of A is

$$A = BCE = ABCDF = DEF$$

The complete alias relationships for this design are in Table 14.25. In general, the resolution of a 2^{k-p} design is equal to the number of letters in the shortest word in the complete defining relation. Therefore, this is a resolution IV design; main effects are aliased with three-factor and higher interactions, and two-factor interactions are aliased with each other. This design would provide good information on the main effects and would give some idea about the strength of the two-factor interactions. The construction and analysis of the design are illustrated in Example 14.9.

EXAMPLE 14.9 | Injection Molding

Parts manufactured in an injection-molding process are showing excessive shrinkage, which is causing problems in assembly operations upstream from the injection-molding area. In an effort to reduce the shrinkage, a quality-improvement team has decided to use a designed experiment to study the injection-molding process. The team investigates each of six factors—mold temperature (A), screw speed (B), holding time (C), cycle time (D), gate size (E), and holding pressure (F)—at two levels with the objective of learning how each factor affects shrinkage and obtaining preliminary information about how the factors interact.

The team decides to use a 16-run two-level fractional factorial design for these six factors. The design is constructed by writing a 2^4 as the basic design in the factors A, B, C, and D and then setting $E = ABC$ and $F = BCD$ as discussed previously. Table 14.26 shows the design along with the observed shrinkage (×10) for the test part produced at each of the 16 runs in the design.

TABLE 14.26 A 2_{IV}^{6-2} Design for the Injection-Molding Experiment

Run	*A*	*B*	*C*	*D*	*E* = *ABC*	*F* = *BCD*	Observed Shrinkage (×10)
1	−	−	−	−	−	−	6
2	+	−	−	−	+	−	10
3	−	+	−	−	+	+	32
4	+	+	−	−	−	+	60
5	−	−	+	−	+	+	4
6	+	−	+	−	−	+	15
7	−	+	+	−	−	−	26
8	+	+	+	−	+	−	60
9	−	−	−	+	−	+	8
10	+	−	−	+	+	+	12
11	−	+	−	+	+	−	34
12	+	+	−	+	−	−	60
13	−	−	+	+	+	−	16
14	+	−	+	+	−	−	5
15	−	+	+	+	−	+	37
16	+	+	+	+	+	+	52

A normal probability plot of the effect estimates from this experiment is shown in Figure 14.34. The only large effects are *A* (mold temperature), *B* (screw speed), and the *AB* interaction. In light of the alias relationship in Table 14.25, it seems reasonable to tentatively adopt these conclusions. The plot of the *AB* interaction in Figure 14.35 shows that the process is insensitive to temperature if the screw speed is at the low level but sensitive to temperature if the screw speed is at the high level. With the screw speed at a low level, the process should produce an average shrinkage of around 10% regardless of the temperature level chosen.

Based on this initial analysis, the team decides to set both the mold temperature and the screw speed at the low level. This set of conditions should reduce the mean shrinkage of parts to around 10%. However, the variability in shrinkage from part to part is still a potential problem. In effect, the mean shrinkage can be adequately reduced by the preceding modifications; however, the part-to-part variability in shrinkage over a production run could still cause problems in assembly. One way to address this issue is to see whether any of the process factors affect the variability in parts shrinkage.

Figure 14.36 presents the normal probability plot of the residuals. This plot appears satisfactory. The plots of residuals versus each factor were then constructed. One of these plots, that for residuals versus factor *C* (holding time), is shown in Figure 14.37. The plot reveals much less scatter in the residuals at the low holding time than at the high holding time. These residuals were obtained in the usual way from a model for predicted shrinkage

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_{12} x_1 x_2 = 27.3125 + 6.9375 x_1 + 17.8125 x_2 + 5.9375 x_1 x_2$$

where x_1, x_2, and x_1x_2 are coded variables that correspond to the factors *A* and *B* and the *AB* interaction. The regression model used to produce the residuals essentially removes the location effects of *A*, *B*, and *AB* from the data; the residuals therefore contain information about unexplained variability. Figure 14.37 indicates that there is a pattern in the variability and that the variability in the shrinkage of parts may be smaller when the holding time is at the low level.

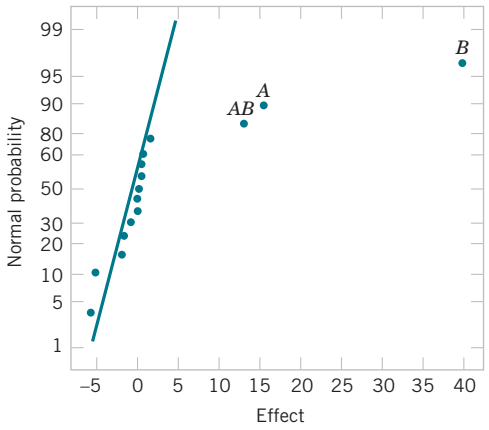


FIGURE 14.34

Normal probability plot of effects for the injection-molding experiment.

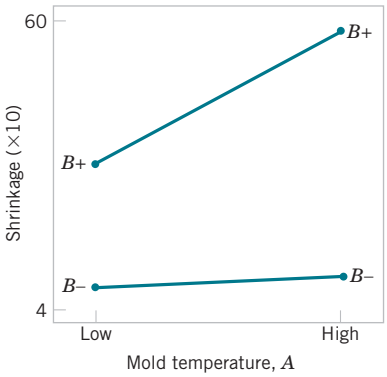


FIGURE 14.35

Plot of *AB* (mold temperature–screw speed) interaction for the injection-molding experiment.

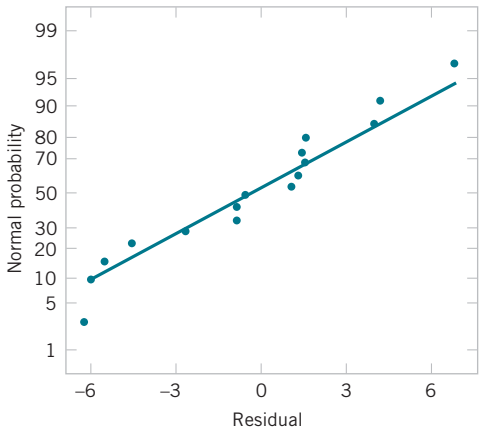


FIGURE 14.36

Normal probability plot of residuals for the injection-molding experiment.

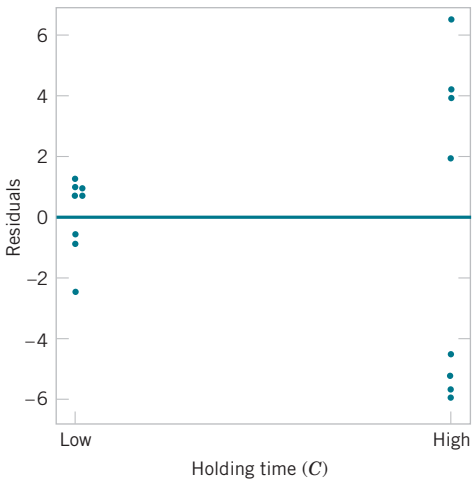


FIGURE 14.37

Residuals versus holding time (*C*) for the injection-molding experiment.

Practical Interpretation: Figure 14.38 shows the data from this experiment projected onto a cube in the factors *A*, *B*, and *C*. The average observed shrinkage and the range of observed shrinkage are shown at each corner of the cube. From inspection of this figure, we see that running the process with the screw speed (*B*) at the low level is the key to reducing average parts shrinkage. If *B* is low, virtually any combination of temperature (*A*) and holding time (*C*) results in low values of average parts shrinkage. However, from examining the ranges of the shrinkage values at each corner of the cube, it is immediately clear that setting the holding time (*C*) at the low level is the most appropriate choice if we wish to keep the part-to-part variability in shrinkage low during a production run.

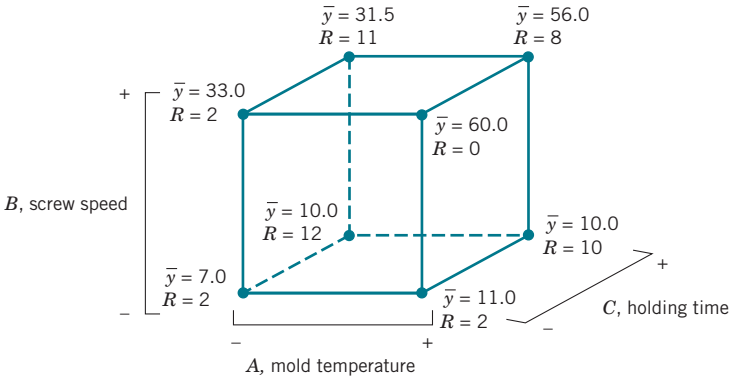


FIGURE 14.38

Average shrinkage and range of shrinkage in factors *A*, *B*, and *C*.

The concepts used in constructing the 2^{6-2} fractional factorial design in Example 14.9 can be extended to the construction of any 2^{k-p} fractional factorial design. In general, a 2^k fractional factorial design containing 2^{k-p} runs is called a $1/2^p$ fraction of the 2^k design or, more simply, a 2^{k-p} fractional factorial design. These designs require the selection of p independent generators. The defining relation for the design consists of the p generators initially chosen and their $2^p - p - 1$ generalized interactions.

The alias structure may be found by multiplying each effect column by the defining relation. Care should be exercised in choosing the generators so that effects of potential interest are not aliased with each other. Each effect has $2^p - 1$ aliases. For moderately large values of k, we usually assume that higher-order interactions (say, third- or fourth-order or higher) to be negligible, and this greatly simplifies the alias structure.

It is important to select the p generators for the 2^{k-p} fractional factorial design in such a way that we obtain the best possible alias relationships. A reasonable criterion is to select the generators so that the resulting 2^{k-p} design has the highest possible design resolution. Montgomery (2017) presented a table of recommended generators (and alias relationships) for 2^{k-p} fractional factorial designs for $k \leq 15$ factors and up to as many as $n \leq 128$ runs. A portion of his table is reproduced here as Table 14.27. In this table, the generators are shown with either + or − choices; selection of all generators as + provides a principal fraction, and if any generators are − choices, the design is one of the alternate fractions for the same family. The suggested generators in this table produce a design of the highest possible resolution.

For example, the 2^{7-4} design is an eight-run experiment accommodating seven variables. This is a 1/16th fraction that is obtained by first writing a 2^3 design as the basic design in the factors A, B, and C, and then forming the four new columns from $I = ABD$, $I = ACE$, $I = BCF$, and $I = ABCG$, as suggested in Table 14.27. The design is shown in Table 14.28.

The complete defining relation is found by multiplying the generators together two, three, and finally four at a time, producing

$$I = ABD = ACE = BCF = ABCG = BCDE = ACDF = CDG = ABEF$$
$$= BEG = AFG = DEF = ADEG = CEFG = BDFG = ABCDEFG$$

The alias of any main effect is found by multiplying that effect through each term in the defining relation. For example, the alias of A is

$$A = BD = CE = ABCF = BCG = ABCDE = CDF = ACDG$$
$$= BEF = ABEG = FG = ADEF = DEG = ACEFG = ABDFG = BCDEFG$$

TABLE 14.27 **Selected 2^{k-p} Fractional Factorial Designs**

Number of Factors k	Fraction	Number of Runs	Design Generators	Number of Factors k	Fraction	Number of Runs	Design Generators
3	2^{3-1}_{III}	4	$C = \pm AB$	9	2^{9-2}_{IV}	128	$H = \pm ACDFG$
4	2^{4-1}_{IV}	8	$D = \pm ABC$				$J = \pm BCEFG$
5	2^{5-1}_{V}	16	$E = \pm ABCD$		2^{9-3}_{IV}	64	$G = \pm ABCD$
	2^{5-2}_{III}	8	$D = \pm AB$				$H = \pm ACEF$
			$E = \pm AC$				$J = \pm CDEF$
6	2^{6-1}_{VI}	32	$F = \pm ABCDE$		2^{9-4}_{IV}	32	$F = \pm BCDE$
	2^{6-2}_{IV}	16	$E = \pm ABC$				$G = \pm ACDE$
			$F = \pm BCD$				$H = \pm ABDE$
	2^{6-3}_{III}	8	$D = \pm AB$				$J = \pm ABCE$
			$E = \pm AC$		2^{9-5}_{III}	16	$E = \pm ABC$
			$F = \pm BC$				$F = \pm BCD$
7	2^{7-1}_{VII}	64	$G = \pm ABCDEF$				$G = \pm ACD$
	2^{7-2}_{IV}	32	$F = \pm ABCD$				$H = \pm ABD$
			$G = \pm ABDE$				$J = \pm ABCD$
	2^{7-3}_{IV}	16	$E = \pm ABC$	10			$H = \pm ABCG$
			$F = \pm BCD$				$J = \pm ACDE$
			$G = \pm ACD$		2^{10-3}_{V}	128	$K = \pm ACDF$
	2^{7-4}_{III}	8	$D = \pm AB$				$G = \pm BCDF$
			$E = \pm AC$				$H = \pm ACDF$
			$F = \pm BC$				$J = \pm ABDE$
			$G = \pm ABC$		2^{10-4}_{IV}	64	$K = \pm ABCE$
8	2^{8-2}_{V}	64	$G = \pm ABCD$				$F = \pm ABCD$
			$H = \pm ABEF$				$G = \pm ABCE$
	2^{8-3}_{IV}	32	$F = \pm ABC$				$H = \pm ABDE$
			$G = \pm ABD$				$J = \pm ACDE$
			$H = \pm BCDE$		2^{10-5}_{IV}	32	$K = \pm BCDE$
	2^{8-4}_{IV}	16	$E = \pm BCD$				$E = \pm ABC$
			$F = \pm ACD$				$F = \pm BCD$
			$G = \pm ABC$				$G = \pm ACD$
			$H = \pm ABD$				$H = \pm ABD$
							$J = \pm ABCD$
					2^{10-6}_{III}	16	$K = \pm AB$

TABLE 14.27 (*Continued*)

Number of Factors k	Fraction	Number of Runs	Design Generators	Number of Factors k	Fraction	Number of Runs	Design Generators
11			$G = \pm CDE$				$K = \pm ADE$
			$H = \pm ABCD$		2_{IV}^{11-6}	32	$L = \pm BDE$
			$J = \pm ABF$				$E = \pm ABC$
			$K = \pm BDEF$				$F = \pm BCD$
	2_{IV}^{11-5}	64	$L = \pm ADEF$				$G = \pm ACD$
			$F = \pm ABC$				$H = \pm ABD$
			$G = \pm BCD$				$J = \pm ABCD$
			$H = \pm CDE$				$K = \pm AB$
			$J = \pm ACD$		2_{III}^{11-7}	16	$L = \pm AC$

Source: Montgomery (2012).

This design is of resolution III because the main effect is aliased with two-factor interactions. If we assume that all three-factor and higher interactions are negligible, the aliases of the seven main effects are

$$\ell_A = A + BD + CE + FG$$
$$\ell_B = B + AD + CF + EG$$
$$\ell_C = C + AE + BF + DG$$
$$\ell_D = D + AB + CG + EF$$
$$\ell_E = E + AC + BG + DF$$
$$\ell_F = F + BC + AG + DE$$
$$\ell_G = G + CD + BE + AF$$

This 2_{III}^{7-4} design is called a *saturated fractional factorial* because all the available degrees of freedom are used to estimate main effects. It is possible to combine sequences of these

TABLE 14.28 A 2_{III}^{7-3} Fractional Factorial Design

A	B	C	$D = AB$	$E = AC$	$F = BC$	$G = ABC$
−	−	−	+	+	+	−
+	−	−	−	−	+	+
−	+	−	−	+	−	+
+	+	−	+	−	−	−
−	−	+	+	−	−	+
+	−	+	−	+	−	−
−	+	+	−	−	+	−
+	+	+	+	+	+	+

resolution III fractional factorials to separate the main effects from the two-factor interactions. The procedure is illustrated in Montgomery (2012).

14.11 Response Surface Methods and Designs

Response surface methodology, or RSM, is a collection of mathematical and statistical techniques that are useful for modeling and analysis in applications in which a response of interest is influenced by several variables and the objective is to **optimize this response**. For example, suppose that a chemical engineer wishes to find the levels of temperature (x_1) and feed concentration (x_2) that maximize the yield (y) of a process. The process yield is a function of the levels of temperature and feed concentration, say,

$$Y = f(x_1, x_2) + \epsilon$$

where ϵ represents the noise or error observed in the response Y. If we denote the expected response by $E(Y) = f(x_1, x_2) = \eta$, then the surface represented by

$$\eta = f(x_1, x_2)$$

is called a **response surface**.

We may represent the response surface graphically as shown in Figure 14.39, where η is plotted versus the levels of x_1 and x_2. Notice that the response is represented as a surface plot in a three-dimensional space. To help visualize the shape of a response surface, we often plot the contours of the response surface as shown in Figure 14.40. In the contour plot, lines of constant response are drawn in the x_1, x_2 plane. Each contour corresponds to a particular height of the response surface. The contour plot is helpful in studying the levels of x_1 and x_2 that result in changes in the shape or height of the response surface.

In most RSM problems, the form of the relationship between the response and the independent variables is unknown. Thus, the first step in RSM is to find a suitable approximation for the true relationship between Y and the independent variables. Usually, a low-order polynomial in some region of the independent variables is employed. If the response is well modeled by a linear function of the independent variables, the approximating function is the **first-order model**

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \tag{14.16}$$

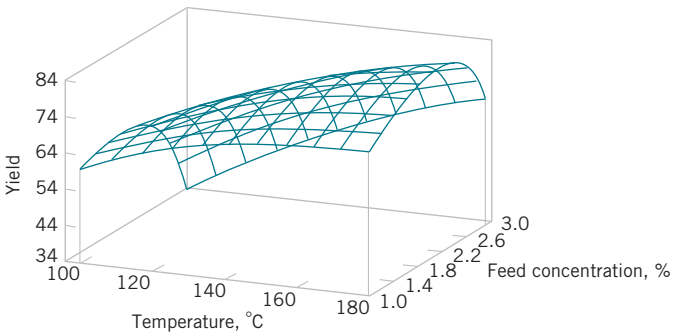


FIGURE 14.39

A three-dimensional response surface showing the expected yield as a function of temperature and feed concentration.

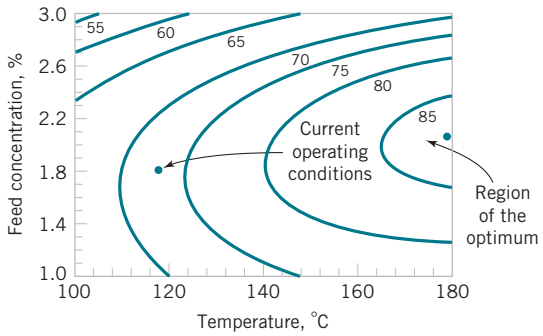


FIGURE 14.40

A contour plot of the yield response surface in Figure 14.39.

If there is curvature in the system, then a polynomial of higher degree must be used, such as the **second-order model**

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum \sum_{i<j} \beta_{ij} x_i x_j + \epsilon \tag{14.17}$$

Many RSM problems use one or both of these approximating polynomials. Of course, it is unlikely that a polynomial model will be a reasonable approximation of the true functional relationship over the entire space of the independent variables, but for a relatively small region, they usually work quite well.

The method of least squares, discussed in Chapters 11 and 12, is used to estimate the parameters in the approximating polynomials. The response surface analysis is then based on the fitted surface. If the fitted surface is an adequate approximation of the true response function, analysis of the fitted surface will be approximately equivalent to analysis of the actual system.

RSM is a **sequential** procedure. Often, when we are at a point on the response surface that is remote from the optimum, such as the current operating conditions in Figure 14.40, there is little curvature in the system and the first-order model is appropriate. Our objective here is to lead the experimenter rapidly and efficiently to the general vicinity of the optimum. Once the region of the optimum has been found, a more elaborate model such as the second-order model may be employed, and an analysis may be performed to locate the optimum. From Figure 14.40, we see that the analysis of a response surface can be thought of as "climbing a hill," where the top of the hill represents the point of maximum response. If the true optimum is a point of minimum response, we may think of "descending into a valley."

The eventual objective of RSM is to determine the optimum operating conditions for the system or to determine a region of the factor space in which operating specifications are satisfied. Also note that the word *optimum* in RSM is used in a special sense. The "hill-climbing" procedures of RSM guarantee convergence to a local optimum only.

Method of Steepest Ascent Frequently, the initial estimate of the optimum operating conditions for the system is far from the actual optimum. In such circumstances, the objective of the experimenter is to move rapidly to the general vicinity of the optimum. We wish to use a simple and economically efficient experimental procedure. When we are remote from the optimum, we usually assume that a first-order model is an adequate approximation to the true surface in a small region of the x's.

The method of **steepest ascent** is a procedure for moving sequentially along the path of steepest ascent, that is, in the direction of the maximum increase in the response. Of course, if **minimization** is desired, we are talking about the method of **steepest descent**. The fitted first-order model is

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i \tag{14.18}$$

and the first-order response surface, that is, the contours of $\hat{y}$, is a series of parallel lines such as that shown in Figure 14.41. The direction of steepest ascent is the direction in which $\hat{y}$ increases most rapidly. This direction is normal to the fitted response surface contours. We usually take as the path of steepest ascent the line through the center of the region of interest and normal to the fitted surface contours. Thus, the steps along the path are proportional to the regression coefficients $\{\hat{\beta}_i\}$. The experimenter determines the actual step size based on process knowledge or other practical considerations.

Experiments are conducted along the path of steepest ascent until no further increase in response is observed. Then a new first-order model may be fit, a new direction of steepest ascent determined, and further experiments conducted in that direction until the experimenter believes that the process is near the optimum.

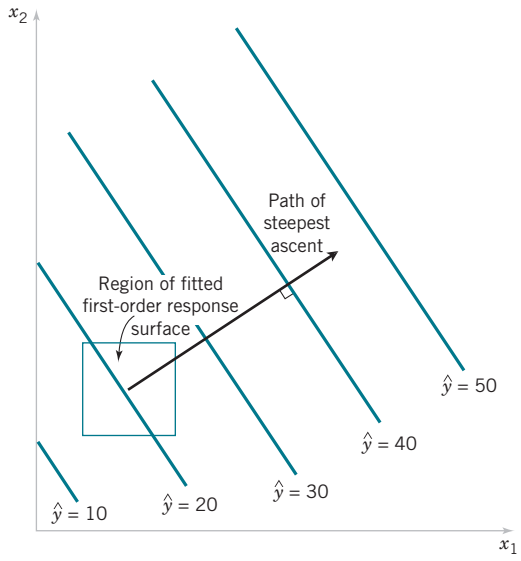


FIGURE 14.41

First-order response surface and path of steepest ascent.

EXAMPLE 14.10 | Process Yield Steepest Ascent

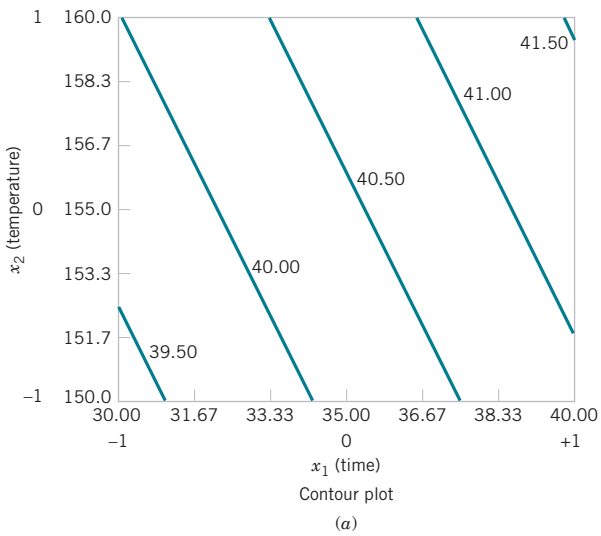
In Example 14.6, we described an experiment on a chemical process in which two factors, reaction time (x_1) and reaction temperature (x_2), affect the percent conversion or yield (Y). Figure 14.24 shows the 2^2 design plus five center points used in this study. The engineer found that both factors were important, there was no interaction, and there was no curvature in the response surface. Therefore, the first-order model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

should be appropriate. Now the effect estimate of time is 1.55 hours and the effect estimate of temperature is 0.65°F, and because the regression coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ are one-half of the corresponding effect estimates, the fitted first-order model is

$$\hat{y} = 40.44 + 0.775x_1 + 0.325x_2$$

Figure 14.42(a) and (b) show the contour plot and three-dimensional surface plot of this model. Figure 14.42 also shows the relationship between the **coded variables** x_1



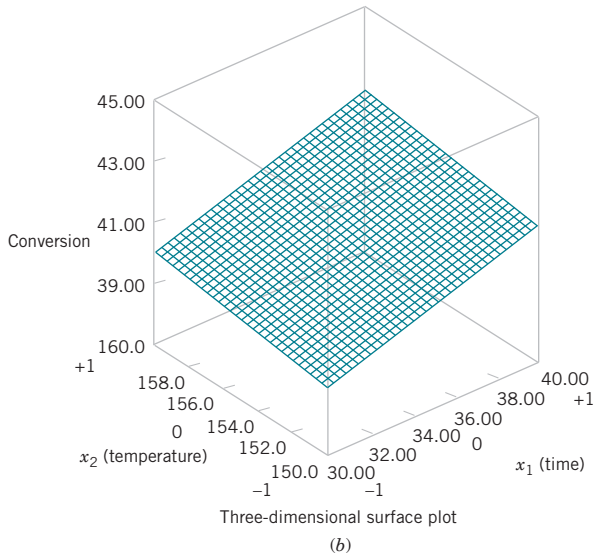


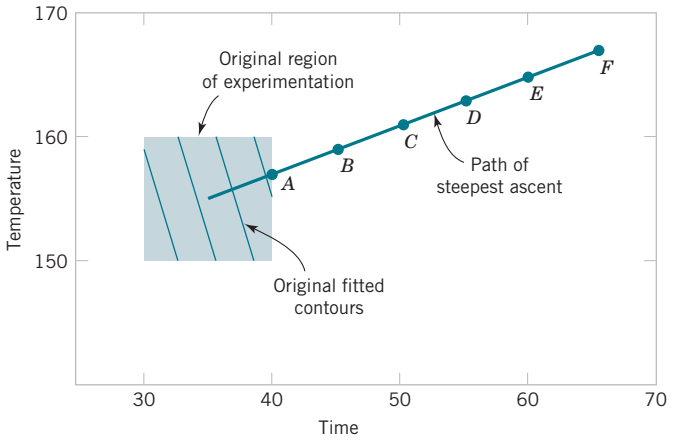
FIGURE 14.42

Response surface plots for the first-order model.

and x_2 (that defined the high and low levels of the factors) and the original variables, time (in minutes) and temperature (in °F).

From examining these plots (or the fitted model), we see that to move away from the design center—the point ($x_1 = 0$, $x_2 = 0$)—along the path of steepest ascent, we would move 0.775 unit in the x_1 direction for every 0.325 unit in the x_2 direction. Thus, the path of steepest ascent passes through the point ($x_1 = 0$, $x_2 = 0$) and has a slope 0.325/0.775. The engineer decides to use 5 minutes of reaction time as the basic step size. Now, 5 minutes of reaction time is equivalent to a step in the *coded* variable x_1 of $\Delta x_1 = 1$. Therefore, the steps along the path of steepest ascent are $\Delta x_1 = 1.0000$ and $\Delta x_2 = (0.325/0.775)\Delta x_1 = 0.42$. A change of $\Delta x_2 = 0.42$ in the coded variable x_2 is equivalent to about 2°F in the original variable temperature. Therefore, the engineer moves along the path of steepest ascent by increasing reaction time by 5 minutes and temperature by 2°F. An actual observation on yield is determined at each point.

Next Steps: Figure 14.43 shows several points along this path of steepest ascent and the yields actually observed from the process at those points. At points A–D, the observed yield increases steadily, but beyond point D, the yield decreases. Therefore, steepest ascent terminates in the vicinity of 55 minutes of reaction time and 163°F with an observed percent conversion of 67%.



Point A: 40 minutes, 157 °F, $y = 40.5$
Point B: 45 minutes, 159 °F, $y = 51.3$
Point C: 50 minutes, 161 °F, $y = 59.6$
Point D: 55 minutes, 163 °F, $y = 67.1$
Point E: 60 minutes, 165 °F, $y = 63.6$
Point F: 65 minutes, 167 °F, $y = 60.7$

FIGURE 14.43

Steepest ascent experiment.

Analysis of a Second-Order Response Surface

When the experimenter is relatively close to the optimum, a second-order model is usually required to approximate the response because of curvature in the true response surface. The fitted second-order model is

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i + \sum_{i=1}^{k} \hat{\beta}_{ii} x_i^2 + \sum_{i<j} \sum \hat{\beta}_{ij} x_i x_j$$

where $\hat{\beta}$ denotes the least squares estimate of b. In this section, we show how to use this fitted model to find the optimum set of operating conditions for the x's and to characterize the nature of the response surface.

EXAMPLE 14.11 | Process Yield Central Composite Design

Continuation of Example 14.10

Consider the chemical process from Example 14.10, where the method of steepest ascent terminated at a reaction time of 55 minutes and a temperature of 163°F. The experimenter decides to fit a second-order model in this region. Table 14.29 and Figure 14.44 show the experimental design, which consists of a 2^2 design centered at 55 minutes and 165°F, five center points, and four runs along the coordinate axes called *axial runs*. This type of design, called a central composite design, is a very popular design for fitting second-order response surfaces.

TABLE 14.29 Central Composite Design

Observation Number	Time (minutes)	Temperature (°F)	Coded Variables x_1	Coded Variables x_2	Conversion (percent) Response 1	Viscosity (mPa-sec) Response 2
1	50	160	−1	−1	65.3	35
2	60	160	1	−1	68.2	39
3	50	170	−1	1	66.0	36
4	60	170	1	1	69.8	43
5	48	165	−1.414	0	64.5	30
6	62	165	1.414	0	69.0	44
7	55	158	0	−1.414	64.0	31
8	55	172	0	1.414	68.5	45
9	55	165	0	0	68.9	37
10	55	165	0	0	69.7	34
11	55	165	0	0	68.5	35
12	55	165	0	0	69.4	36
13	55	165	0	0	69.0	37

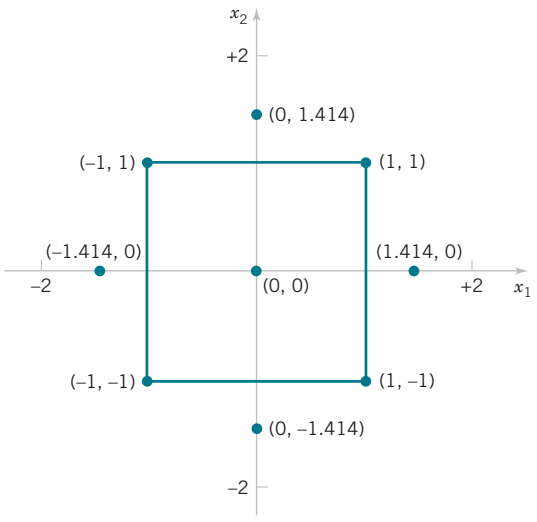


FIGURE 14.44
Central composite design.

Two response variables were measured during this phase of the experiment: percentage conversion (yield) and viscosity. The least-squares quadratic model for the yield response is

$$\hat{y} = 69.1 + 1.633x_1 + 1.083x_2 - 0.969x_1^2 - 1.219x_2^2 + 0.225x_1x_2$$

The analysis of variance for this model is shown in Table 14.30.

Figure 14.45 shows the response surface contour plot and the three-dimensional surface plot for this model. From examination of these plots, the maximum yield is about 70%, obtained at approximately 60 minutes of reaction time and 167°F.

The viscosity response is adequately described by the first-order model

$$\hat{y}_2 = 37.08 + 3.85x_1 + 3.10x_2$$

Table 14.31 summarizes the analysis of variance for this model. The response surface is shown graphically in Figure 14.46. Notice that viscosity increases as both time and temperature increase.

TABLE 14.30 Analysis of Variance for the Quadratic Model, Yield Response

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	*P*-Value
Model	45.89	5	9.178	14.93	0.0013
Residual	4.30	7	0.615		
Total	50.19	12			

Independent Variable	Coefficient	SE Coefficient	t	*P*-Value
Intercept	69.100	0.351	197.1	
x_1	1.633	0.277	5.891	0.0006
x_2	1.083	0.277	3.907	0.0058
x_1^2	−0.969	0.297	−3.259	0.0139
x_2^2	−1.219	0.297	−4.100	0.0046
x_1x_2	0.225	0.392	0.574	0.5839

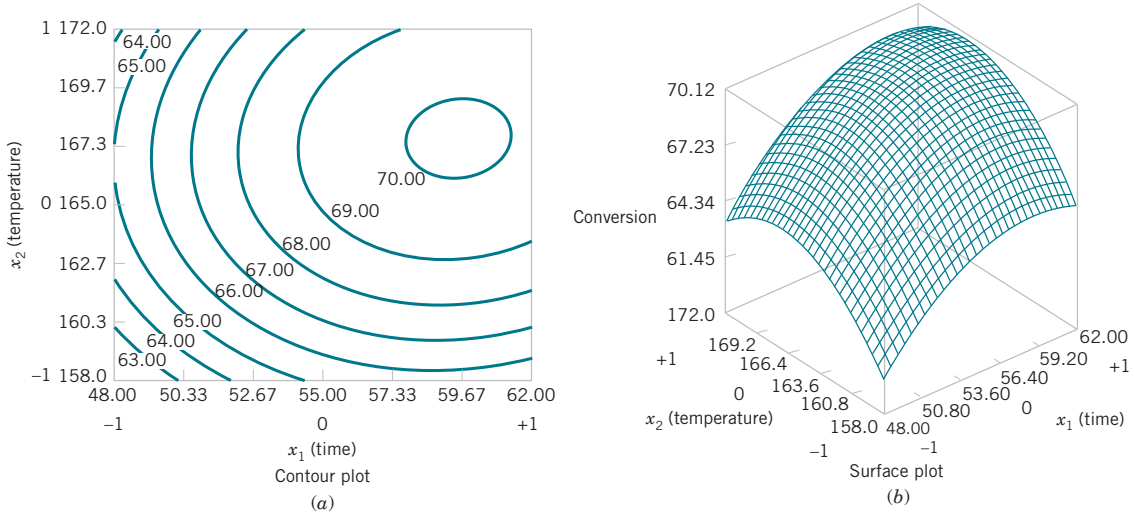


FIGURE 14.45
Response surface plots for the yield response.

TABLE 14.31 Analysis of Variance for the First-Order Model, Viscosity Response

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	*P*-Value
Model	195.4	2	97.72	15.89	0.0008
Residual	61.5	10	6.15		
Total	256.9	12			

Independent Variable	Coefficient	SE Coefficient	t	*P*-Value
Intercept	37.08	0.69	53.910	
x_1	3.85	0.88	4.391	0.0014
x_2	3.10	0.88	3.536	0.0054

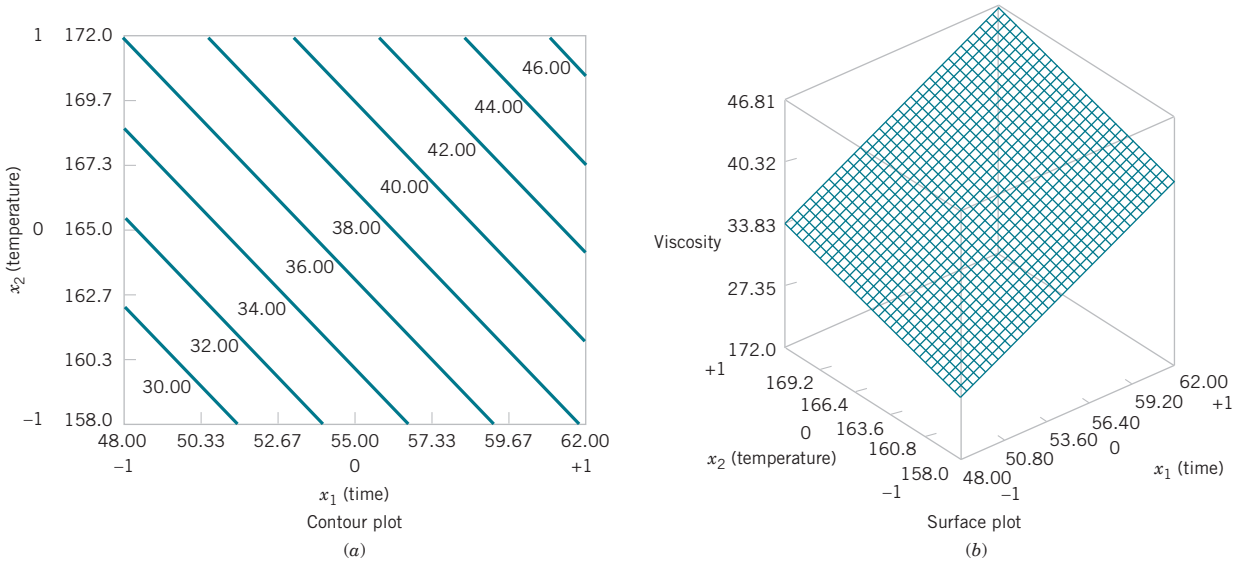


FIGURE 14.46

Response surface plots for the viscosity response.

Practical Interpretation: As in most response surface problems, the experimenter in this example had conflicting objectives regarding the two responses. The objective was to maximize yield, but the acceptable range for viscosity was $38 \leq y_2 \leq 42$. When there are only a few independent variables, an easy way to solve this problem is to overlay the response surfaces to find the optimum. Figure 14.47 shows the overlay plot of both responses with the contours $y_1 = 69\%$ conversion, $y_2 = 38$, and $y_2 = 42$ highlighted. The shaded areas on this plot identify unfeasible combinations of time and temperature. This graph shows that several combinations of time and temperature are satisfactory.

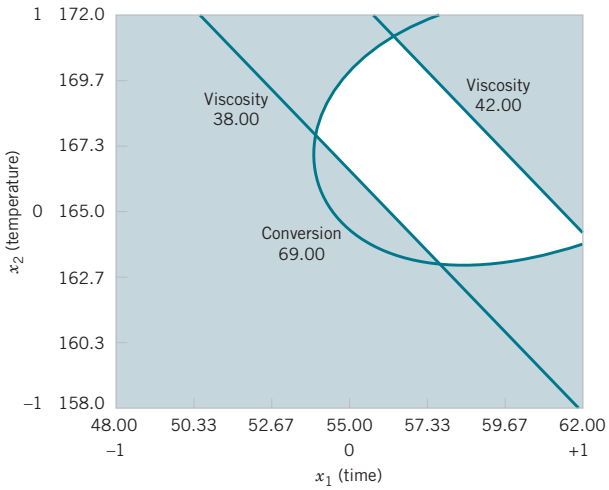


FIGURE 14.47

Overlay of yield and viscosity response surfaces.

Example 14.11 illustrates the use of a **central composite design** (CCD) for fitting a second-order response surface model. These designs are widely used in practice because they are relatively efficient with respect to the number of runs required. In general, a CCD in k factors requires 2^k factorial runs, $2k$ axial runs, and at least one center point (three to five center points are typically used). Designs for $k = 2$ and $k = 3$ factors are shown in Figure 14.48.

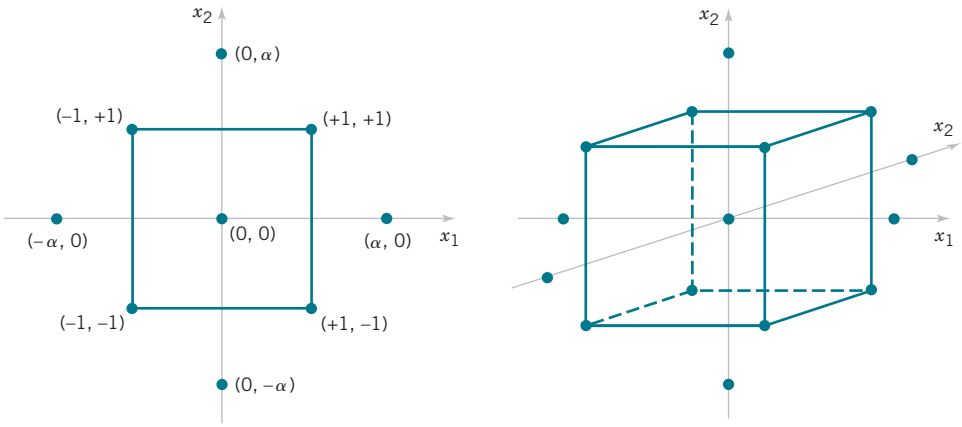


FIGURE 14.48

Central composite designs for $k = 2$ and $k = 3$.

The central composite design may be made **rotatable** by proper choice of the axial spacing α in Figure 14.48. If the design is rotatable, the standard deviation of predicted response $\hat{y}$ is constant at all points that are the same distance from the center of the design. For rotatability, choose $a = (F)^{1/4}$ where F is the number of points in the factorial part of the design (usually $F = 2^k$). For the case of $k = 2$ factors, $\alpha = (2^2)^{1/4} = 1.414$, as was used in the design in Example 14.11. Figure 14.49 presents a contour plot and a surface plot of the standard deviation of prediction for the quadratic model used for the yield response. Notice that the contours are concentric circles, implying that yield is predicted with equal precision for all points that are the same distance from the center of the design. Also, as one would expect, the precision decreases with increasing distance from the design center.

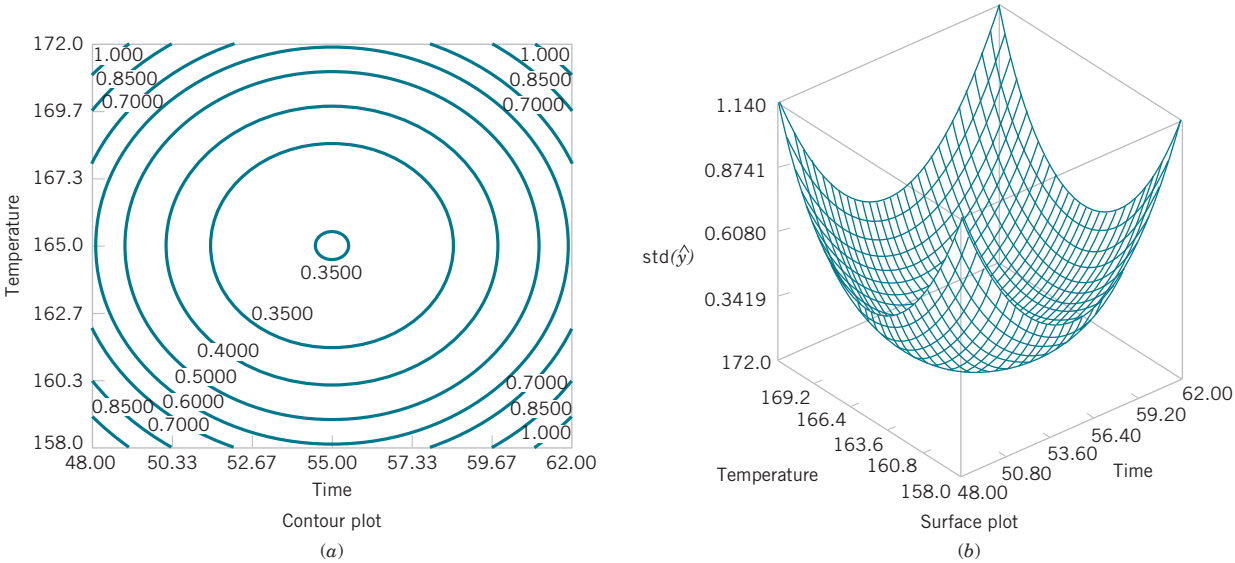


FIGURE 14.49

Plots of constant $\sqrt{V(\hat{y})}$ for a rotatable central composite design.

Important Terms and Concepts

Aliases
Analysis of variance (ANOVA)
Blocking and nuisance factors
Center points
Central composite design
Complete defining relation
Confounding
Contrast
Defining relation
Design matrix
Factorial experiment
Fractional factorial design
Generator
Interaction
Main effect
Normal probability plot of factor effects
Optimization experiment
Orthogonal design
Projection property
Regression model
Residual analysis
Resolution
Response surface
Screening experiment
Steepest ascent (or descent)
2^k factorial design (Two-level factorial design)

CHAPTER **15**

Statistical Quality Control

doc-stock/Alamy Stock Photo

Courtesy of Mike Johnson, www.redbead.com

LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Understand the role of statistical tools in quality improvement
2. Understand the different types of variability, rational subgroups, and use of a control chart to detect assignable causes
3. Understand the general form of a Shewhart control chart and how to apply zone rules (such as the Western Electric rules) and pattern analysis to detect assignable causes
4. Construct and interpret control charts for variables such as $\bar{X}$, R, S, and individuals charts
5. Construct and interpret control charts for attributes such as P and U charts
6. Calculate and interpret process capability ratios
7. Calculate the ARL performance for a Shewhart control chart
8. Construct and interpret cumulative sum and exponentially weighted moving-average control charts
9. Use other statistical process control problem-solving tools

CHAPTER OUTLINE

15.1 Quality Improvement and Statistics
- 15.1.1 Statistical Quality Control
- 15.1.2 Statistical Process Control

15.2 Introduction to Control Charts
- 15.2.1 Basic Principles
- 15.2.2 Design of a Control Chart
- 15.2.3 Rational Subgroups
- 15.2.4 Analysis of Patterns on Control Charts

15.3 $\bar{X}$ and R or S Control Charts

15.4 Control Charts for Individual Measurements

15.5 Process Capability

15.6 Attribute Control Charts
- 15.6.1 P Chart (Control Chart for Proportions)
- 15.6.2 U Chart (Control Chart for Defects per Unit)

15.7 Control Chart Performance

15.8 Time-Weighted Charts
- 15.8.1 Exponentially Weighted Moving-Average Control Chart
- 15.8.2 Cumulative Sum Control Chart

15.9 Other SPC Problem-Solving Tools

15.10 Decision Theory
- 15.10.1 Decision Models
- 15.10.2 Decision Criteria

15.11 Implementing SPC

Bowl of Beads

Quality guru W. Edwards Deming conducted a simple experiment in his seminars with a bowl of beads. Many were colored white but a percentage of red beads was randomly mixed in the bowl. A participant from the seminar was provided a paddle made with indentations so that 50 beads at a time could be scooped from the bowl. The participant was allowed to use only the paddle and instructed to scoop only white beads (repeated multiple times with beads replaced). The red beads were considered to be defectives. Of course, this was difficult to do, and each scoop resulted in a count of red beads. Deming plotted the fraction of red beads from each scoop and used the results to make several points. As was clear from the scenario, this process was beyond the participant's ability to make simple improvements. The process needed to be changed (reduce the number of red beads), which is the responsibility of management. Furthermore, many business processes have this type of characteristic, and it is important to learn from the data whether the variability is common, intrinsic to the process, or is the result of some special cause. This distinction is important for the type of process control or improvements to be applied. Refer to the example of control adjustments in Chapter 1. Control charts are primary tools to understand process variability, and that is the main topic of this chapter.

15.1 Quality Improvement and Statistics

The quality of products and services is a major decision factor in most businesses. Regardless of whether the consumer is an individual, a corporation, a military defense program, or a retail store, a consumer making purchase decisions is likely to consider quality of equal importance to cost and schedule. Consequently, **quality improvement** is a major concern to many U.S. corporations.

Quality means **fitness for use**. For example, we purchase automobiles that we expect to be free of manufacturing defects and that should provide reliable and economical transportation or a retailer buys finished goods with the expectation that they are properly packaged and arranged for easy storage and display. In other words, all customers expect that the products and services they buy meet their requirements. Those requirements define fitness for use.

Quality or fitness for use is determined through the interaction of **quality of design** and **quality of conformance**. By quality of design, we mean the different grades or levels of performance, reliability, serviceability, and function that are the result of deliberate engineering and management decisions. By quality of conformance, we mean the systematic **reduction of variability** and **elimination of defects** until every unit produced is identical and defect free.

Quality improvement means better quality of design through improved knowledge of customers' requirements and the systematic **elimination of waste**. Examples of waste include scrap and rework in manufacturing, inspection and testing, errors on documents (such as engineering

drawings, checks, purchase orders, and plans), customer complaint hotlines, warranty costs, and the time required to do things over again that could have been done right the first time. A successful quality-improvement effort can eliminate much of this waste and lead to lower costs, higher productivity, increased customer satisfaction, increased business reputation, higher market share, and ultimately higher profits for the company.

Statistical methods play a vital role in quality improvement. In product design and development, statistical methods, including designed experiments, can be used to compare different materials, components, or ingredients, and to help determine both system and component tolerances. Statistical methods can be used to determine the capability of a manufacturing process. Statistical process control can be used to systematically improve a process by reducing variability. Life testing provides reliability and other performance data about the product. This can lead to new and improved designs and products that have longer useful lives and lower operating and maintenance costs. Regression methods are often used to determine key process indicators that link with customer satisfaction. This enables organizations to focus on the most important process measurements. In this chapter, we provide an introduction to the basic methods of statistical quality control that, along with experimental design, form the basis of a successful quality-improvement effort.

15.1.1 Statistical Quality Control

This chapter is about **statistical quality control**, a collection of tools that are essential in quality-improvement activities. The field of statistical quality control can be broadly defined as the use of those statistical and engineering methods in measuring, monitoring, controlling, and improving quality. Statistical quality control is a field that dates back to the 1920s. Dr. Walter A. Shewhart of Bell Telephone Laboratories was one of the early pioneers of the field. In 1924, he wrote a memorandum showing a modern control chart, one of the basic tools of statistical process control. Harold F. Dodge and Harry G. Romig, two other Bell System employees, provided much of the leadership in the development of statistically based sampling and inspection methods. The work of these three men forms much of the basis of the modern field of statistical quality control. The widespread introduction of these methods to U.S. industry occurred during World War II. Dr. W. Edwards Deming and Dr. Joseph M. Juran were instrumental in spreading statistical quality-control methods since World War II.

The Japanese have been particularly successful in deploying statistical quality-control methods and have used such methods to gain significant advantage over their competitors. In the 1970s, U.S. industry suffered extensively from Japanese (and other foreign) competition; that has led, in turn, to renewed interest in statistical quality-control methods in the United States. Much of this interest focuses on *statistical process control* and *experimental design*. Many U.S. companies have implemented these methods in their manufacturing, engineering, and other business organizations.

15.1.2 Statistical Process Control

It is impractical to inspect quality into a product; the product must be built right the first time. The manufacturing process must therefore be stable or repeatable and capable of operating with little variability around the target or nominal dimension. Online statistical process control is a powerful tool for achieving process stability and improving capability through the reduction of variability.

It is customary to think of **statistical process control (SPC)** as a set of **problem-solving tools** that may be applied to any process. The major tools of SPC are histogram, Pareto chart, cause-and-effect diagram, defect-concentration diagram, control chart, scatter diagram, and check sheet. The control chart is the most powerful of the SPC tools.

15.2 Introduction to Control Charts

15.2.1 Basic Principles

In any production process, regardless of how well designed or carefully maintained it is, a certain amount of inherent or natural variability always exists. This natural variability or "background

noise" is the cumulative effect of many small, essentially unavoidable causes. When the background noise in a process is relatively small, we usually consider it an acceptable level of process performance. In the framework of statistical quality control, this natural variability is often called a "stable system of chance causes." A process that is operating with only **chance causes** of variation present is said to be in *statistical control*. In other words, the chance causes are an inherent part of the process.

Other kinds of variability may occasionally be present in the output of a process. This variability in key quality characteristics usually arises from three sources: improperly adjusted machines, operator errors, or defective raw materials. Such variability is generally large when compared to the background noise, and it usually represents an unacceptable level of process performance. We refer to these sources of variability that are not part of the chance cause pattern as **assignable causes**. A process that is operating in the presence of assignable causes is said to be *out of control*.*

Production processes often operate in the in-control state, producing acceptable product for relatively long periods of time. Occasionally, however, assignable causes occur, seemingly at random, resulting in a "shift" to an out-of-control state in which a large proportion of the process output does not conform to requirements. A major objective of statistical process control is to quickly detect the occurrence of assignable causes or process shifts so that investigation of the process and corrective action can be undertaken before many nonconforming units are manufactured. The **control chart** is an online process-monitoring technique widely used for this purpose.

Recall the following from Chapter 1. Adjustments to common causes of variation increase the variation of a process whereas actions should be taken in response to assignable causes of variation. Control charts may also be used to estimate the parameters of a production process and, through this information, to determine the capability of a process to meet specifications. The control chart can also provide information that is useful in improving the process. Finally, remember that the eventual goal of statistical process control is the *elimination of variability in the process*. Although it may not be possible to eliminate variability completely, the control chart helps reduce it as much as possible.

A typical control chart is shown in Figure 15.1, which is a graphical display of a quality characteristic that has been measured or computed from a sample versus the sample number or time. Often, the samples are selected at periodic intervals such as every few minutes or every hour.

- The chart contains a **center line** (CL) that represents the average value of the quality characteristic corresponding to the in-control state. (That is, only chance causes are present.)
- Two other horizontal lines, called the *upper control limit* (*UCL*) and the *lower control limit* (*LCL*), are also shown on the chart. These **control limits** are chosen so that if the process is in control, nearly all of the sample points fall between them.

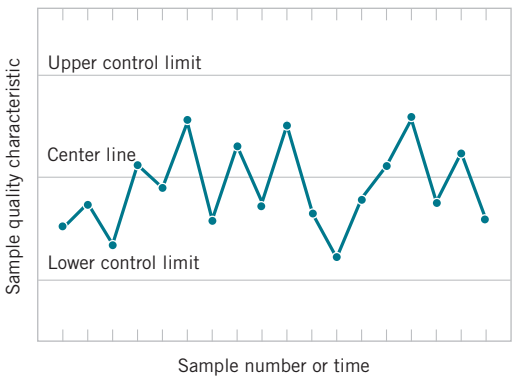


FIGURE 15.1
A typical control chart.

*Dr. Walter A. Shewhart developed the terminology *chance* and *assignable* causes. Today some writers use *common* cause instead of *chance* cause and *special* cause instead of *assignable* cause.

In general, as long as the points plot within the control limits, the process is assumed to be in control, and no action is necessary. However, a point that plots outside of the control limits is interpreted as evidence that the process is out of control, and investigation and corrective action are required to find and eliminate the assignable cause or causes responsible for this behavior. The sample points on the control chart are usually connected with straight-line segments so that it is easier to visualize how the sequence of points has evolved over time.

Even if all the points plot inside the control limits, if they behave in a systematic or nonrandom manner, this is an indication that the process is out of control. For example, if 18 of the last 20 points plotted above the center line but below the upper control limit, and only two of these points plotted below the center line but above the lower control limit, we would be very suspicious that something was wrong. If the process is in control, all plotted points should have an essentially random pattern. Methods designed to find sequences or nonrandom patterns can be applied to control charts as an aid in detecting out-of-control conditions. A particular nonrandom pattern usually appears on a control chart for a reason, and if that reason can be found and eliminated, process performance can be improved.

A close connection exists between control charts and hypothesis testing. Essentially, the control chart is a series of tests of the hypothesis that the process is in a state of statistical control. A point plotting within the control limits is equivalent to failing to reject the hypothesis of statistical control, and a point plotting outside the control limits is equivalent to rejecting the hypothesis of statistical control.

We give a general *model* for a control chart. Let W be a sample statistic that measures some quality characteristic of interest, and suppose that the mean of W is μ_W and the standard deviation of W is σ_W.[†] Then the center line, the upper control limit, and the lower control limit become

Control Chart Model

$$\begin{aligned} UCL &= \mu_W + k\sigma_W \\ CL &= \mu_W \\ LCL &= \mu_W - k\sigma_W \end{aligned} \tag{15.1}$$

where k is the "distance" of the control limits from the center line expressed in standard deviation units. A common choice is $k = 3$. Dr. Walter A. Shewhart first proposed this general theory of control charts, and those developed according to these principles are often called **Shewhart control charts**.

The control chart is a device for describing exactly what statistical control means; as such, it may be used in a variety of ways. In many applications, the control chart is used for online process monitoring. That is, sample data are collected and used to construct the control chart, and if the sample values of $\bar{x}$ (say) fall within the control limits and do not exhibit any systematic pattern, we say the process is in control at the level indicated by the chart. Note that we may be interested here in determining *both* whether the past data came from a process that was in control and whether future samples from this process indicate statistical control.

The most important use of a control chart is to *improve* the process. Most processes do not operate in a state of statistical control. Consequently, the routine and attentive use of control charts identifies assignable causes. If these causes can be eliminated from the process, variability is reduced and the process is improved. This process-improvement activity using the control chart is illustrated in Figure 15.2. Notice that the control chart only *detects* assignable causes. Management, operator, and engineering *action* usually are necessary to eliminate the assignable cause. An action plan for responding to control chart signals is vital. In identifying and

[†]Note that "sigma" refers to the standard deviation of the statistic plotted on the chart (i.e., σ_W), not the standard deviation of the quality characteristic.

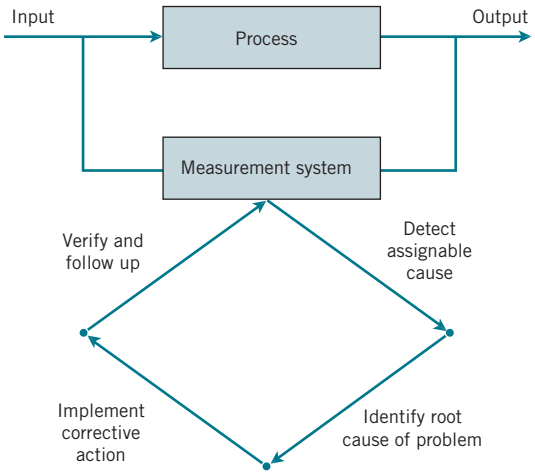


FIGURE 15.2
Process improvement using the control chart.

eliminating assignable causes, it is important to find the underlying **root cause** of the problem and to attack it. A cosmetic solution does not result in any real, long-term process improvement. Developing an effective system for corrective action is an essential component of an effective SPC implementation.

We may also use the control chart as an *estimating device*. That is, from a control chart that exhibits statistical control, we may estimate certain process parameters, such as the mean, standard deviation, and fraction nonconforming or fallout. These estimates may then be used to determine the *capability* of the process to produce acceptable products. Such **process capability** studies have considerable impact on many management decision problems that occur over the product cycle, including make-or-buy decisions, plant and process improvements that reduce process variability, and contractual agreements with customers or suppliers regarding product quality. Such estimates are discussed in a later section.

Control charts may be classified into two general types.

- Many quality characteristics can be measured and expressed as numbers on some continuous scale of measurement. In such cases, it is convenient to describe the quality characteristic with a measure of central tendency and a measure of variability. Control charts for central tendency and variability are collectively called **variables control charts**. The $\overline{X}$ chart is the most widely used chart for monitoring central tendency, and charts based on either the sample range or the sample standard deviation are used to control process variability.
- Many quality characteristics are not measured on a continuous scale or even a quantitative scale. In these cases, we may judge each unit of product as either conforming or nonconforming on the basis of whether or not it possesses certain attributes, or we may count the number of nonconformities (defects) appearing on a unit of product. Control charts for such quality characteristics are called **attributes control charts**.

Control charts have had a long history of use in industry. There are at least five reasons for their popularity:

1. **Control charts are a proven technique for improving productivity.** A successful control chart program reduces scrap and rework, which are the primary productivity killers in *any* operation. If you reduce scrap and rework, productivity increases, cost decreases, and production capacity (measured in the number of *good* parts per hour) increases.
2. **Control charts are effective in defect prevention.** The control chart helps keep the process in control, which is consistent with the "do it right the first time" philosophy. It is never cheaper to sort out the "good" units from the "bad" later on than it is to build them correctly initially. If you do not have effective process control, you are paying someone to make a nonconforming product.

3. **Control charts prevent unnecessary process adjustments.** A control chart can distinguish between background noise and abnormal variation; no other device, including a human operator, is as effective in making this distinction. If process operators adjust the process based on periodic tests unrelated to a control chart program, they often overreact to the background noise and make unneeded adjustments. These unnecessary adjustments can result in a deterioration of process performance. In other words, the control chart is consistent with the "if it isn't broken, don't fix it" philosophy.
4. **Control charts provide diagnostic information.** Frequently, the pattern of points on the control chart contains information that is of diagnostic value to an experienced operator or engineer. This information allows the operator to implement a change in the process that improve its performance.
5. **Control charts provide information about process capability.** The control chart provides information about the value of important process parameters and their stability over time. This allows an estimate of process capability to be made. This information is of tremendous use to product and process designers.

Control charts are among the most effective management control tools, and they are as important as cost controls and material controls. Modern computer technology has made it easy to implement control charts in any type of process because data collection and analysis can be performed in real time, online at the work center.

15.2.2 Design of a Control Chart

To illustrate these ideas, we give a simplified example of a control chart. In manufacturing automobile engine piston rings, the inside diameter of the rings is a critical quality characteristic. The process mean inside ring diameter is 74 millimeters, and it is known that the standard deviation of ring diameter is 0.01 millimeters. A control chart for average ring diameter is shown in Figure 15.3. Every few minutes a random sample of five rings is taken, the average ring diameter of the sample (say, $\bar{x}$) is computed, and $\bar{x}$ is plotted on the chart. Because this control chart utilizes the sample mean $\overline{X}$ to monitor the process mean, it is usually called an $\overline{X}$ control chart. Note that all the points fall within the control limits, so the chart indicates that the process is in statistical control.

Consider how the control limits were determined. The process average is 74 millimeters, and the process standard deviation is $\sigma = 0.01$ millimeters. Now if samples of size $n = 5$ are taken, the standard deviation of the sample average $\overline{X}$ is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.01}{\sqrt{5}} = 0.0045$$

Therefore, if the process is in control with a mean diameter of 74 millimeters, by using the central limit theorem to assume that $\overline{X}$ is approximately normally distributed, we would expect approximately $100(1 - \alpha)\%$ of the sample mean diameters $\overline{X}$ to fall between $74 + z_{\alpha/2}(0.0045)$ and

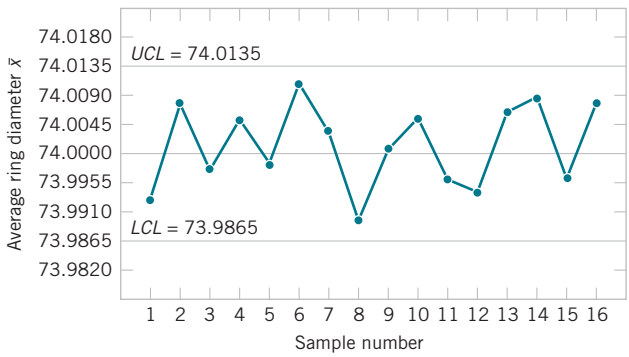


FIGURE 15.3

$\overline{X}$ control chart for piston ring diameter.

$74 - z_{\alpha/2}(0.0045)$. As discussed previously, we customarily choose the constant $z_{\alpha/2}$ to be 3, so the upper and lower control limits become

$$UCL = 74 + 3(0.0045) = 74.0135$$

and

$$LCL = 74 - 3(0.0045) = 73.9865$$

as shown on the control chart. These are the 3-sigma control limits referred to earlier. Note that the use of 3-sigma limits implies that $\alpha = 0.0027$; that is, the probability that the point plots outside the control limits when the process is in control is 0.0027. The width of the control limits is inversely related to the sample size n for a given multiple of sigma. Choosing the control limits is equivalent to setting up the critical region for testing the hypotheses

$$H_0: \mu = 74 \qquad H_1: \mu \neq 74$$

where $\sigma = 0.01$ is known. Essentially, the control chart tests this hypothesis repeatedly at different points in time.

In designing a control chart, we must specify both the sample size to use and the frequency of sampling. In general, larger samples make it easier to detect small shifts in the process.

We must also determine the frequency of sampling. The most desirable situation from the view of detecting shifts would be to take large samples very frequently; however, this is usually not economically feasible. The general problem is one of *allocating sampling effort*. That is, either we take small samples at short intervals or larger samples at longer intervals. Current industry practice tends to favor smaller, more frequent samples, particularly in high-volume manufacturing processes or when a great many types of assignable causes can occur.

Furthermore, as automatic sensing and measurement technology develops, it is becoming possible to greatly increase frequencies. Ultimately, every unit can be tested as it is manufactured. This capability does not eliminate the need for control charts because the test system cannot prevent defects. The increased data expand the effectiveness of process control and improve quality.

When preliminary samples are used to construct limits for control charts, these limits are customarily treated as trial values. Therefore, the sample statistics should be plotted on the appropriate charts, and any points that exceed the control limits should be investigated. If assignable causes for these points are discovered, they should be eliminated and new limits for the control charts determined. In this way, the process may be eventually brought into statistical control and its inherent capabilities assessed. Other changes in process centering and dispersion may then be contemplated.

15.2.3 Rational Subgroups

A fundamental idea in the use of control charts is to collect sample data according to what Shewhart called the **rational subgroup** concept. Generally, this means that subgroups or samples should be selected so that to the extent possible, the variability of the observations within a subgroup should include all the chance or natural variability and exclude the assignable variability. Then, the control limits represent bounds for all the chance variability, not the assignable variability. Consequently, assignable causes tend to generate points that are outside of the control limits, and chance variability tends to generate points that are within the control limits.

When control charts are applied to production processes, the time order of production is a logical basis for rational subgrouping. Even though time order is preserved, it is still possible to form subgroups erroneously. If some of the observations in the subgroup are taken at the end of one 8-hour shift and the remaining observations are taken at the start of the next 8-hour shift, any differences between shifts are handled as chance variability when, instead, it should be considered as assignable variability. This makes detecting differences between shifts more difficult. Still, in general, time order is frequently a good basis for forming subgroups because it allows us to detect assignable causes that occur over time.

In the most common approach to rational subgroups, each subgroup consists of units that were produced at the same time (or as closely together as possible). This approach is used when the primary purpose of the control chart is to detect process shifts. It minimizes variability due to assignable causes *within* a sample, and it maximizes variability *between* samples if assignable causes are present. It also provides estimates of the standard deviation of the process in the case of variables control charts. This approach to rational subgrouping essentially gives a "snapshot" of the process at each point in time when a sample is collected.

Other bases for forming rational subgroups can be used. For example, suppose that a process consists of several machines that pool their output into a common stream. If we sample from this common stream of output, it is very difficult to detect whether or not some of the machines are out of control. A logical approach to rational subgrouping here is to apply control chart techniques to the output for each individual machine. Sometimes this concept needs to be applied to different heads on the same machine, different workstations, different operators, and so forth.

The rational subgroup concept is very important. The proper selection of samples requires careful consideration of the process with the objective of obtaining as much useful information as possible from the control chart analysis.

15.2.4 Analysis of Patterns on Control Charts

A control chart may indicate an out-of-control condition either when one or more points fall beyond the control limits, or when the plotted points exhibit some nonrandom pattern of behavior. For example, consider the $\overline{X}$ chart shown in Figure 15.4. Although all 25 points fall within the control limits, the points do not indicate statistical control because their pattern is very nonrandom in appearance. Specifically, we note that 19 of the 25 points plot below the center line, but only 6 of them plot above. If the points are truly random, we should expect a more even distribution of points above and below the center line. We also observe that following the fourth point, five points in a row increase in magnitude. This arrangement of points is called a **run**. Because the observations are increasing, we could call it a *run up*; similarly, a sequence of decreasing points is called a *run down*. This control chart has an unusually long run up (beginning with the fourth point) and an unusually long run down (beginning with the 18th point).

In general, we define a *run* as a sequence of observations of the same type. In addition to runs up and runs down, we could define the types of observations as those above and below the center line, respectively, so two points in a row above the center line would be a run of length 2.

A run of length 8 or more points has a very low probability of occurrence in a random sample of points. Consequently, any type of run of length 8 or more is often taken as a signal of an out-of-control condition. For example, 8 consecutive points on one side of the center line indicate that the process is out of control.

Although runs are an important measure of nonrandom behavior on a control chart, other types of patterns may also indicate an out-of-control condition. For example, consider the $\overline{X}$ chart in Figure 15.5. Note that the plotted sample averages exhibit a cyclic behavior, yet they all fall

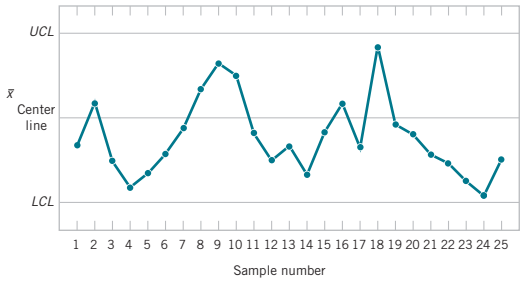


FIGURE 15.4

An $\overline{X}$ control chart.

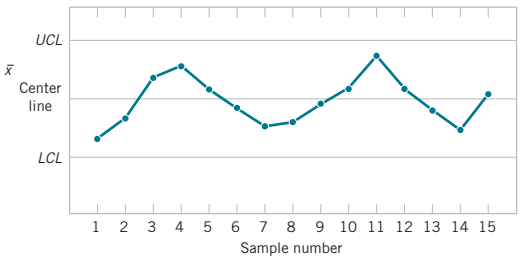


FIGURE 15.5

An $\overline{X}$ chart with a cyclic pattern.

within the control limits. Such a pattern may indicate a problem with the process, such as operator fatigue, raw material deliveries, and heat or stress buildup. The yield may be improved by eliminating or reducing the sources of variability causing this cyclic behavior (Figure 15.6). In Figure 15.6, *LSL* and *USL* denote the lower and upper **specification limits** of the process, respectively. These limits represent bounds within which acceptable product must fall, and they are often based on customer requirements.

The problem is one of **pattern recognition**, that is, recognizing systematic or nonrandom patterns on the control chart and identifying the reason for this behavior. The ability to interpret a particular pattern in terms of assignable causes requires experience and knowledge of the process. That is, we must not only know the statistical principles of control charts, but we must also have a good understanding of the process.

The *Western Electric Handbook* (1956) suggests a set of decision rules for detecting nonrandom patterns on control charts. Specifically, the widely adopted **Western Electric rules** conclude that the process is out of control if either

1. One point plots outside 3-sigma control limits.
2. Two of three consecutive points plot beyond a 2-sigma limit.
3. Four of five consecutive points plot at a distance of 1 sigma or beyond from the center line.
4. Eight consecutive points plot on one side of the center line.

We have found these rules very effective in practice for enhancing the sensitivity of control charts. Rules 2 and 3 apply to one side of the center line at a time. That is, a point above the *upper* 2-sigma limit followed immediately by a point below the *lower* 2-sigma limit would not signal an out-of-control alarm.

Figure 15.7 shows an $\overline{X}$ control chart for the piston ring process with the 1-sigma, 2-sigma, and 3-sigma limits used in the Western Electric procedure. Notice that these inner limits (sometimes called **warning limits**) partition the control chart into three zones A, B, and C on each side of the center line. Consequently, the Western Electric rules are sometimes called the **run rules** for control charts. Notice that the last four points fall in zone B or beyond. Thus, because four of five consecutive points exceed the 1-sigma limit, the Western Electric procedure concludes that the pattern is nonrandom and the process is out of control.

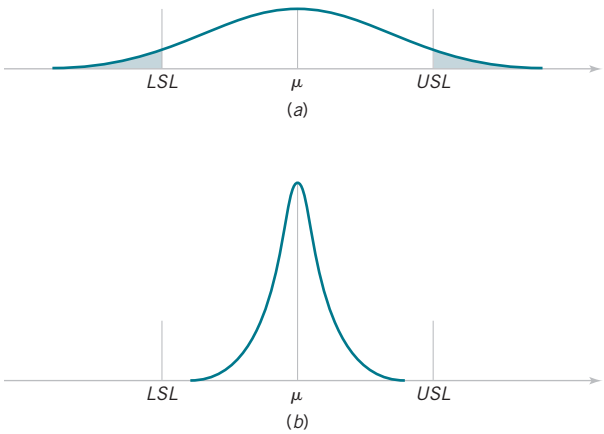


FIGURE 15.6

(a) Variability with the cyclic pattern. (b) Variability with the cyclic pattern eliminated.

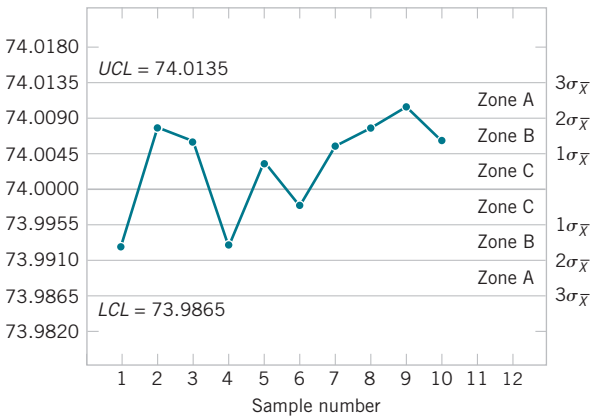


FIGURE 15.7

The Western Electric zone rules.

15.3 $\bar{X}$ and R or S Control Charts

When dealing with a quality characteristic that can be expressed as a measurement, monitoring both the mean value of the quality characteristic and its variability is customary. Control over the average quality is exercised by the control chart for averages, usually called the $\bar{X}$ chart. Process variability can be controlled by either a range chart (***R* chart**) or a standard deviation chart (***S* chart**), depending on how the population standard deviation is estimated.

We usually estimate parameters on the basis of preliminary samples taken when the process is thought to be in control. We recommend the use of at least 20 to 25 preliminary samples. Suppose that m preliminary samples are available, each of size n. Typically, n is 4, 5, or 6; these relatively small sample sizes are widely used and often arise from the construction of rational subgroups. Let the sample mean for the ith sample be $\bar{X}_i$. Then we estimate the mean of the population, μ, by the **grand mean**

$$\hat{\mu} = \bar{\bar{X}} = \frac{1}{m}\sum_{i=1}^{m} \bar{X}_i \tag{15.2}$$

Thus, we may take $\bar{\bar{X}}$ as the center line on the $\bar{X}$ control chart.

We may estimate σ from either the standard deviation or the range of the observations within each sample. The sample size is usually relatively small, so there is little loss in efficiency in estimating σ from the sample ranges.

$\bar{X}$ and R Charts The relationship between the range R of a sample from a normal population with known parameters and the standard deviation of that population is needed. Because R is a random variable, the quantity $W = R/\sigma$, called the *relative range*, is also a random variable. The mean and standard deviation of the distribution of W are called d_2 and d_3, respectively. The values for d_2 and d_3 depend on the subgroup size n. They are computed numerically and available in tables or computer software. Because $R = \sigma W$,

$$\mu_R = d_2\sigma \qquad \sigma_R = d_3\sigma \tag{15.3}$$

Let R_i be the range of the ith sample, and let

$$\bar{R} = \frac{1}{m}\sum_{i=1}^{m} R_i \tag{15.4}$$

be the average range. Then $\bar{R}$ is an estimator of μ_R and from Equation 15.3 we obtain the following.

Estimator of σ from *R* Chart

An unbiased estimator of σ is

$$\hat{\sigma} = \frac{\bar{R}}{d_2} \tag{15.5}$$

where the constant d_2 is tabulated for various sample sizes in Appendix Table XI.

Therefore, once we have computed the sample values $\bar{\bar{x}}$ and $\bar{r}$, we may use as our upper and lower control limits for the $\bar{X}$ chart

$$\begin{aligned} UCL &= \hat{\mu}_{\bar{X}} + 3\hat{\sigma}_{\bar{X}} = \hat{\mu} + \frac{3\hat{\sigma}}{\sqrt{n}} = \bar{\bar{x}} + \frac{3}{d_2\sqrt{n}}\bar{r} \\ LCL &= \hat{\mu}_{\bar{X}} - 3\hat{\sigma}_{\bar{X}} = \hat{\mu} - \frac{3\hat{\sigma}}{\sqrt{n}} = \bar{\bar{x}} - \frac{3}{d_2\sqrt{n}}\bar{r} \end{aligned} \tag{15.6}$$

Define the constant

$$A_2 = \frac{3}{d_2\sqrt{n}} \tag{15.7}$$

Now, the $\overline{X}$ control chart may be defined as follows.

$\overline{X}$ Control Chart (from $\overline{R}$)

The center line and upper and lower control limits for an $\overline{X}$ control chart are

$$UCL = \bar{\bar{x}} + A_2\,\bar{r} \qquad CL = \bar{\bar{x}} \qquad LCL = \bar{\bar{x}} - A_2\,\bar{r} \tag{15.8}$$

where the constant A_2 is tabulated for various sample sizes in Appendix Table XI.

The parameters of the R chart may also be determined easily. The center line is $\overline{R}$. To determine the control limits, we need an estimate of σ_R, the standard deviation of R. Once again, assuming that the process is in control, the distribution of the relative range, W, is useful. We may estimate σ_R from Equation 15.3 as

$$\hat{\sigma}_R = d_3\hat{\sigma} = d_3\frac{\overline{R}}{d_2} \tag{15.9}$$

and the upper and lower control limits on the R chart are

$$\begin{aligned} UCL &= \bar{r} + \frac{3d_3}{d_2}\bar{r} = \left(1 + \frac{3d_3}{d_2}\right)\bar{r} \\ LCL &= \bar{r} - \frac{3d_3}{d_2}\bar{r} = \left(1 - \frac{3d_3}{d_2}\right)\bar{r} \end{aligned} \tag{15.10}$$

Setting $D_3 = 1 - 3d_3/d_2$ and $D_4 = 1 + 3d_3/d_2$ leads to the following definition.

R Chart

The center line and upper and lower control limits for an R chart are

$$UCL = D_4\bar{r} \qquad CL = \bar{r} \qquad LCL = D_3\bar{r} \tag{15.11}$$

where $\bar{r}$ is the sample average range, and the constants D_3 and D_4 are tabulated for various sample sizes in Appendix Table XI.

The LCL for an R chart can be a negative number. In that case, it is customary to set LCL to zero. Because the points plotted on an R chart are non-negative, no points can fall below an LCL of zero. Also, we often study the R chart first because if the process variability is not constant over time, the control limits calculated for the $\overline{X}$ chart can be misleading.

$\overline{X}$ and S Charts Rather than base control charts on ranges, a more modern approach is to calculate and plot the standard deviation of each subgroup to monitor the process standard deviation σ. This is called an S chart. When an S chart is used, it is common to use these standard deviations to develop control limits for the $\overline{X}$ chart. Typically, the sample size used for subgroups is small (fewer than 10) and in that case there is usually little difference in the $\overline{X}$ chart generated from ranges or standard deviations. However, because computer software is often used to implement control charts, S charts are quite common. Details to construct these charts follow.

Section 7.3 stated that S is a biased estimator of σ. That is, $E(S) = c_4\sigma$ where c_4 is a constant that is near, but not equal to, 1. Furthermore, a calculation similar to the one used for $E(S)$ can derive the standard deviation of the statistic S with the result $\sigma\sqrt{1-c_4^2}$. Therefore, the center line and 3-sigma control limits for S are

$$LCL = c_4\sigma - 3\sigma\sqrt{1-c_4^2} \qquad CL = c_4\sigma \qquad UCL = c_4\sigma + 3\sigma\sqrt{1-c_4^2} \tag{15.12}$$

Assume that there are m preliminary samples available, each of size n, and let S_i denote the standard deviation of the ith sample. Define

$$\overline{S} = \frac{1}{m}\sum_{i=1}^{m} S_i \tag{15.13}$$

Because $E(\overline{S}) = c_4\sigma$, we obtain the following.

Estimator of σ from S Chart

An unbiased estimator of σ

$$\hat{\sigma} = \overline{S}/c_4 \tag{15.14}$$

where the constant c_4 is tabulated for various sample sizes in Appendix Table XI.

When an S chart is used, the estimate for σ in Equation 15.14 is commonly used to calculate the control limits for an $\overline{X}$ chart. This produces the following control limits for an $\overline{X}$ chart.

$\overline{X}$ Control Chart (from $\overline{S}$)

$$UCL = \overline{\overline{x}} + 3\frac{\overline{s}}{c_4\sqrt{n}} \qquad CL = \overline{\overline{x}} \qquad LCL = \overline{\overline{x}} - 3\frac{\overline{s}}{c_4\sqrt{n}} \tag{15.15}$$

A control chart for standard deviations follows.

S Chart

$$UCL = \overline{s} + 3\frac{\overline{s}}{c_4}\sqrt{1-c_4^2} \qquad CL = \overline{s} \qquad LCL = \overline{s} - 3\frac{\overline{s}}{c_4}\sqrt{1-c_4^2} \tag{15.16}$$

The LCL for an S chart can be a negative number; in that case, it is customary to set LCL to zero.

EXAMPLE 15.1 | Vane Opening

A component part for a jet aircraft engine is manufactured by an investment casting process. The vane opening on this casting is an important functional parameter of the part. We illustrate the use of $\overline{X}, R$, and S control charts to assess the statistical stability of this process. See Table 15.1 for 20 samples of five parts each. The values given in the table have been coded by using the last three digits of the dimension; that is, 31.6 indicates 0.50316 inch.

The quantities $\overline{\overline{x}} = 33.3$ and $\overline{r} = 5.8$ appear at the foot of Table 15.1. The value of A_2 for samples of size 5 is $A_2 = 0.577$ from Appendix Table XI. Then the trial control limits for the $\overline{X}$ chart are

$$\overline{\overline{x}} \pm A_2\,\overline{r} = 33.32 \pm (0.577)(5.8) = 33.32 \pm 3.35$$

or

$$UCL = 36.67 \qquad LCL = 29.97$$

TABLE 15.1 Vane-Opening Measurements

Sample Number	x_1	x_2	x_3	x_4	x_5	$\bar{x}$	r	s
1	33	29	31	32	33	31.6	4	1.67332
2	33	31	35	37	31	33.4	6	2.60768
3	35	37	33	34	36	35.0	4	1.58114
4	30	31	33	34	33	32.2	4	1.64317
5	33	34	35	33	34	33.8	2	0.83666
6	38	37	39	40	38	38.4	3	1.14018
7	30	31	32	34	31	31.6	4	1.51658
8	29	39	38	39	39	36.8	10	4.38178
9	28	33	35	36	43	35.0	15	5.43139
10	38	33	32	35	32	34.0	6	2.54951
11	28	30	28	32	31	29.8	4	1.78885
12	31	35	35	35	34	34.0	4	1.73205
13	27	32	34	35	37	33.0	10	3.80789
14	33	33	35	37	36	34.8	4	1.78885
15	35	37	32	35	39	35.6	7	2.60768
16	33	33	27	31	30	30.8	6	2.48998
17	35	34	34	30	32	33.0	5	2.00000
18	32	33	30	30	33	31.6	3	1.51658
19	25	27	34	27	28	28.2	9	3.42053
20	35	35	36	33	30	33.8	6	2.38747
						$\bar{\bar{x}} = 33.32$	$\bar{r} = 5.8$	$\bar{s} = 2.345$

For the R chart, the trial control limits are

$$UCL = D_4\bar{r} = (2.115)(5.8) = 12.27$$
$$LCL = D_3\bar{r} = (0)(5.8) = 0$$

The $\overline{X}$ and R control charts with these trial control limits are shown in Figure 15.8. Notice that samples 6, 8, 11, and 19 are out of control on the $\overline{X}$ chart and that sample 9 is out of control on the R chart. (These points are labeled with a "1" because they violate the first Western Electric rule.)

For the S chart, the value of $c_4 = 0.94$. Therefore,

$$\frac{3\bar{s}}{c_4}\sqrt{1 - c_4^2} = \frac{3(2.345)}{0.94}\sqrt{1 - 0.94^2} = 2.553$$

and the trial control limits are

$$UCL = 2.345 + 2.553 = 4.898$$
$$LCL = 2.345 - 2.553 = -0.208$$

The LCL is set to zero. If $\bar{s}$ is used to determine the control limits for the $\overline{X}$ chart,

$$\bar{\bar{x}} \pm \frac{3\bar{s}}{c_4\sqrt{n}} = 33.32 \pm \frac{3(2.345)}{0.94\sqrt{5}} = 33.32 \pm 3.35$$

and this result is nearly the same as from $\bar{r}$. The S chart is shown in Figure 15.9. Because the control limits for the $\overline{X}$ chart calculated from $\bar{s}$ are nearly the same as from $\bar{r}$, the chart is not shown.

Suppose that all of these assignable causes can be traced to a defective tool in the wax-molding area. We should discard these five samples and recompute the limits for the $\overline{X}$ and R charts. These new revised limits for the $\overline{X}$ chart are

$$UCL = \bar{\bar{x}} + A_2\,\bar{r} = 33.21 + (0.577)(5.0) = 36.10$$
$$LCL = \bar{\bar{x}} - A_2\,\bar{r} = 33.21 - (0.577)(5.0) = 30.33$$

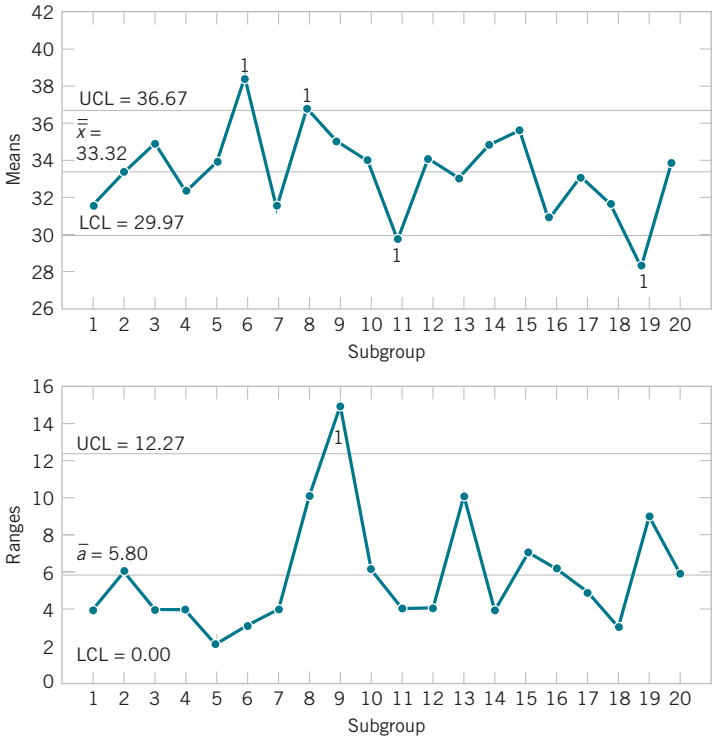


FIGURE 15.8

The $\bar{X}$ and R control charts for vane opening.

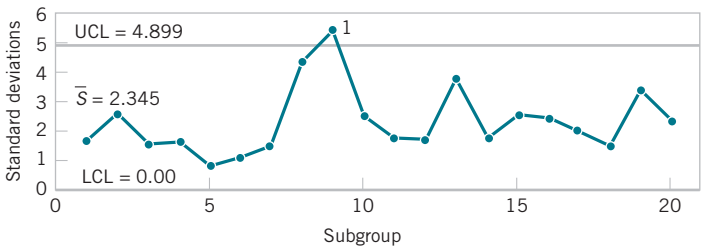


FIGURE 15.9

The S control chart for vane opening.

and for the R chart,

$$UCL = D_4\bar{r} = (2.115)(5.0) = 10.57$$

$$LCL = D_3\bar{r} = (0)(5.0) = 0$$

The revised control charts are shown in Figure 15.10.

Practical Interpretation: Notice that we have treated the first 20 preliminary samples as *estimation data* with which to establish control limits. These limits can now be used to judge the statistical control of future production. As each new sample becomes available, the values of $\bar{x}$ and r should be computed and plotted on the control charts. It may be desirable to revise the limits periodically even if the process remains stable. The limits should always be revised when process improvements are made.

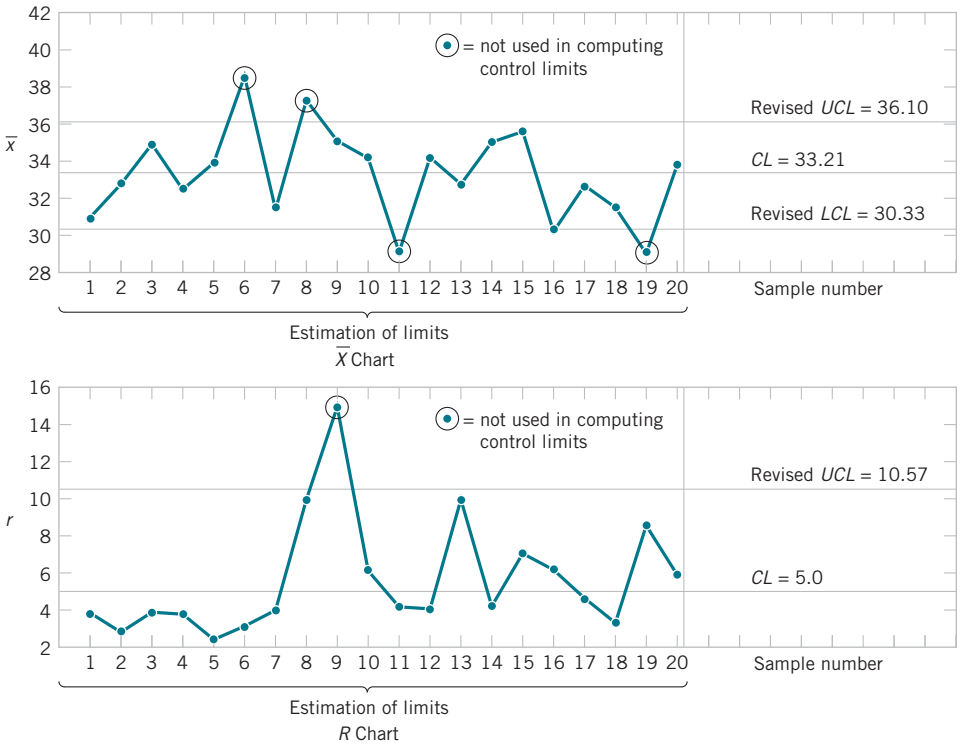


FIGURE 15.10

The $\overline{X}$ and R control charts for vane opening, revised limits.

Computer Construction of $\overline{X}$ and R Control Charts Many computer programs construct $\overline{X}$ and R control charts. Figures 15.8 and 15.10 show charts similar to those produced by computer software for the vane-opening data. Software usually allows the user to select any multiple of sigma as the width of the control limits and use the Western Electric rules to detect out-of-control points. The software also prepares a summary report as in Table 15.2 and excludes subgroups from the calculation of the control limits.

TABLE 15.2 Summary Report from Computer Software for the Vane-Opening Data

Test Results for Xbar Chart
TEST 1. One point more than 3.00 sigmas from center line.
Test Failed at points: 6 8 11 19
Test Results for R Chart
TEST 1. One point more than 3.00 sigmas from center line.
Test Failed at points: 9

15.4 Control Charts for Individual Measurements

In many situations, the sample size used for process control is $n = 1$; that is, the sample consists of an individual unit. Some examples of these situations follow:

1. Automated inspection and measurement technology is used, and every unit manufactured is analyzed.
2. The production rate is very slow, and it is inconvenient to allow sample sizes of $n > 1$ to accumulate before being analyzed.
3. Repeat measurements on the process differ only because of laboratory or analysis error as in many chemical processes.
4. In process plants, such as papermaking, measurements on some parameters such as coating thickness *across* the roll differ very little and produce a standard deviation that is much too small if the objective is to control coating thickness *along* the roll.

In such situations, the **individuals control chart** (also called an ***X* chart**) is useful. The control chart for individuals uses the **moving range** of two successive observations to estimate the process variability. The moving range is defined as $MR_i = |X_i - X_{i-1}|$ and for m observations the average moving range is

$$\overline{MR} = \frac{1}{m-1}\sum_{i=2}^{m} |X_i - X_{i-1}|$$

An estimator of σ is

$$\hat{\sigma} = \frac{\overline{MR}}{d_2} = \frac{\overline{MR}}{1.128} \tag{15.17}$$

where the value for d_2 corresponds to $n = 2$ because each moving range is the range between two consecutive observations. Note that there are only $m - 1$ moving ranges. It is also possible to establish a control chart on the moving range using D_3 and D_4 for $n = 2$. The parameters for these charts are defined as follows.

Individuals Control Chart

The center line and upper and lower control limits for a control chart for individuals are

$$\begin{aligned} UCL &= \bar{x} + 3\frac{\overline{mr}}{d_2} = \bar{x} + 3\frac{\overline{mr}}{1.128} \\ CL &= \bar{x} \\ LCL &= \bar{x} - 3\frac{\overline{mr}}{d_2} = \bar{x} - 3\frac{\overline{mr}}{1.128} \end{aligned} \tag{15.18}$$

and for a control chart for moving ranges

$$\begin{aligned} UCL &= D_4\overline{mr} = 3.267\overline{mr} \\ CL &= \overline{mr} \\ LCL &= D_3\overline{mr} = 0 \end{aligned}$$

Note that LCL for this moving range chart is always zero because $D_3 = 0$ for $n = 2$. The procedure is illustrated in the following example.

EXAMPLE 15.2 | Chemical Process Concentration

Table 15.3 has 20 observations on concentration for the output of a chemical process. The observations are taken at one-hour intervals. If several observations are taken at the same time, the observed concentration readings differ only because of measurement error. Because the measurement error is small, only one observation is taken each hour.

To set up the control chart for individuals, note that the sample average of the 20 concentration readings is $\bar{x} = 99.1$ and that the average of the moving ranges of two observations shown in the last column of Table 15.3 is $\overline{mr} = 2.59$. To set up the moving-range chart, we note that $D_3 = 0$ and $D_4 = 3.267$ for $n = 2$. Therefore, the moving-range chart has center line $\overline{mr} = 2.59, LCL = 0$, and $UCL = D_4\overline{mr} = (3.267)(2.59) = 8.46$. The control chart is shown in Figure 15.11, which was constructed by computer software. Because no points exceed the upper control limit, we may now set up the control chart for individual concentration measurements. If a moving range of $n = 2$ observations is used, $d_2 = 1.128$. For the data in Table 15.3, we have

$$UCL = \bar{x} + 3\frac{\overline{mr}}{d_2} = 99.1 + 3\frac{2.59}{1.128} = 105.99$$

$$CL = \bar{x} = 99.1$$

$$LCL = \bar{x} - 3\frac{\overline{mr}}{d_2} = 99.1 - 3\frac{2.59}{1.128} = 92.21$$

The control charts for individual concentration measurements and moving ranges are shown in Figure 15.11. There is no indication of an out-of-control condition.

Practical Interpretation: These calculated control limits are used to monitor future production.

TABLE 15.3 Chemical Process Concentration Measurements

Observation	Concentration x	Moving Range mr	Observation	Concentration x	Moving Range mr
1	102.0		12	98.7	2.6
2	94.8	7.2	13	101.1	2.4
3	98.3	3.5	14	98.4	2.7
4	98.4	0.1	15	97.0	1.4
5	102.0	3.6	16	96.7	0.3
6	98.5	3.5	17	100.3	3.6
7	99.0	0.5	18	101.4	1.1
8	97.7	1.3	19	97.2	4.2
9	100.0	2.3	20	101.0	3.8
10	98.1	1.9		$\bar{x} = 99.1$	$\overline{mr} = 2.59$
11	101.3	3.2			

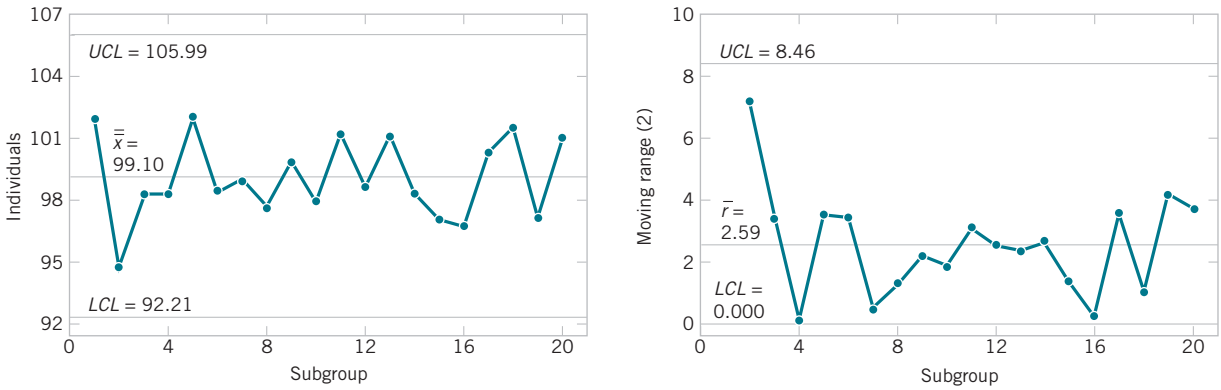


FIGURE 15.11

Control charts for individuals and the moving range from computer software for the chemical process concentration data.

The chart for individuals can be interpreted much like an ordinary $\overline{X}$ control chart. A shift in the process average results in either a point (or points) outside the control limits or a pattern consisting of a run on one side of the center line.

Some care should be exercised in interpreting patterns on the moving-range chart. The moving ranges are correlated, and this correlation may often induce a pattern of runs or cycles on the chart. The individual measurements are assumed to be uncorrelated, however, and any apparent pattern on the individuals' control chart should be carefully investigated.

The control chart for individuals is not very sensitive to small shifts in the process mean. For example, if the size of the shift in the mean is 1 standard deviation, the average number of points to detect this shift is 43.9. This result is shown later in the chapter. Although the performance of the control chart for individuals is much better for large shifts, in many situations the shift of interest is not large and more rapid shift detection is desirable. In these cases, we recommend time-weighted charts such as the *cumulative sum control chart* or an *exponentially weighted moving-average chart* (discussed in Section 15.8).

Some individuals have suggested that limits narrower than 3-sigma be used on the chart for individuals to enhance its ability to detect small process shifts. This is a dangerous suggestion, for narrower limits dramatically increase **false alarms** and the charts may be ignored and become useless. If you are interested in detecting small shifts, consider the time-weighted charts in Section 15.8.

15.5 Process Capability

It is usually necessary to obtain some information about the **process capability**, that is, the performance of the process when it is operating in control. Two graphical tools, the **tolerance chart** (or tier chart) and the **histogram**, are helpful in assessing process capability. The tolerance chart for all 20 samples from the vane-manufacturing process is shown in Figure 15.12. The specifications on vane opening are 0.5030 ± 0.0010 in. In terms of the coded data, the upper specification limit is $USL = 40$ and the lower specification limit is $LSL = 20$, and these limits are shown on the chart in Figure 15.12. Each measurement is plotted on the tolerance chart. Measurements from the same subgroup are connected with lines. The tolerance chart is useful in revealing patterns over time in the individual measurements, or it may show that a particular value of $\bar{x}$ or r was produced by one or two unusual observations in the sample. For example, note the two unusual observations in sample 9 and the single unusual observation in sample 8. Note also that it is appropriate to plot the specification limits on the tolerance chart because it is a chart of individual measurements. **It is never appropriate to plot specification limits on a control chart or to use the specifications in determining the control limits**. Specification limits and control limits are unrelated. Finally, note from Figure 15.12 that the process is running off-center from the nominal dimension of 30 (or 0.5030 inches).

The histogram for the vane-opening measurements is shown in Figure 15.13. The observations from samples 6, 8, 9, 11, and 19 (corresponding to out-of-control points on either the $\overline{X}$ or R chart) have been deleted from this histogram. The general impression from examining this histogram is that the process is capable of meeting the specification but that it is running off-center.

Another way to express process capability is in terms of an index that is defined as follows.

Process Capability Ratio

The **process capability ratio (*PCR*)** is

$$PCR = \frac{USL - LSL}{6\sigma} \tag{15.19}$$

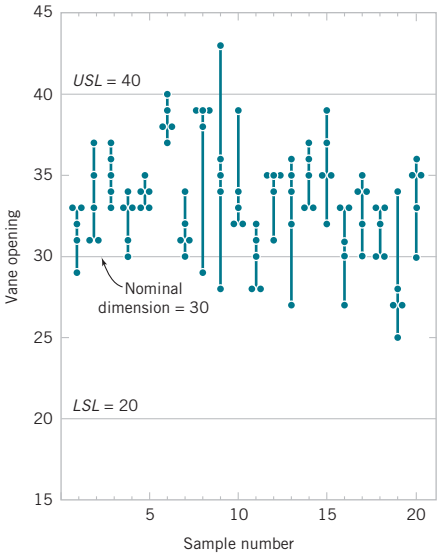


FIGURE 15.12

Tolerance diagram of vane openings.

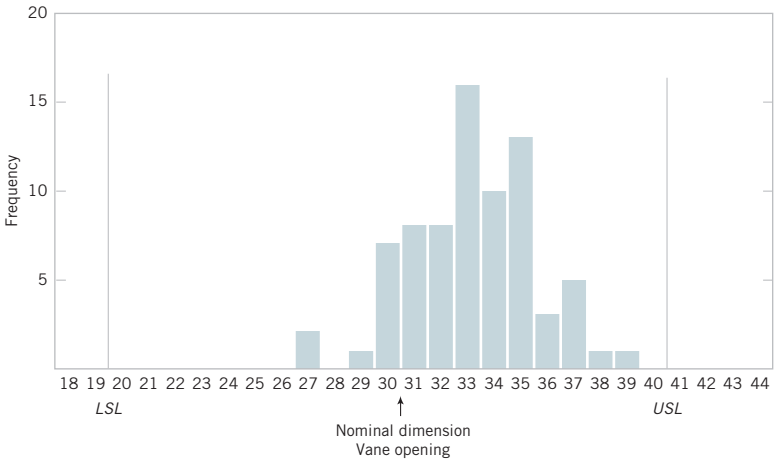


FIGURE 15.13

Histogram for vane opening.

The numerator of *PCR* is the width of the specifications. The limits 3σ on either side of the process mean are sometimes called **natural tolerance limits**, for these represent limits that an in-control process should meet with most of the units produced. Consequently, 6σ is often referred to as the width of the process. For the vane opening, where our sample size is 5, we could estimate σ as

$$\hat{\sigma} = \frac{\bar{r}}{d_2} = \frac{5.0}{2.326} = 2.15$$

Therefore, the *PCR* is estimated to be

$$PCR = \frac{USL - LSL}{6\hat{\sigma}} = \frac{40 - 20}{6(2.15)} = 1.55$$

The *PCR* has a natural interpretation: (1/*PCR*)100% is just the percentage of the specifications' width used by the process. Thus, the vane-opening process uses approximately (1/1.55)100% = 64.5% of the specifications' width.

Figure 15.14(a) shows a process for which the *PCR* exceeds unity. Because the process natural tolerance limits lie inside the specifications, few defective or nonconforming units are produced. If *PCR* = 1, as shown in Figure 15.14(b), more nonconforming units result. In fact, for a normally distributed process, if *PCR* = 1, the fraction nonconforming is 0.27%, or 2700 **parts per million**. Finally, when the *PCR* is less than unity, as in Figure 15.14(c), the process is very yield-sensitive and a large number of nonconforming units is being produced.

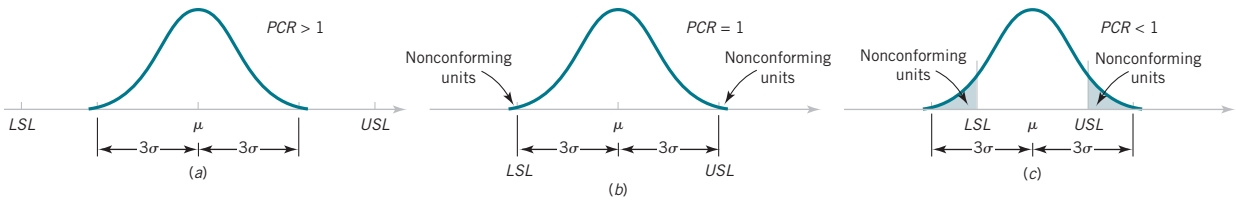


FIGURE 15.14

Process fallout and the process capability ratio (PCR).

The definition of the *PCR* given in Equation 15.19 implicitly assumes that the process is centered at the nominal dimension. If the process is running off-center, its **actual capability** is less than indicated by the *PCR*. It is convenient to think of *PCR* as a measure of **potential capability**, that is, capability with a centered process. If the process is not centered, a measure of actual capability is often used. This ratio, called PCR_k, is defined next.

PCR_k

$$PCR_k = \min\left[\frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right] \tag{15.20}$$

In effect, PCR_k is a one-sided process capability ratio that is calculated relative to the specification limit nearest to the process mean. For the vane-opening process, we find that the estimate of the process capability ratio PCR_k (after deleting the samples corresponding to out-of-control points) is

$$\begin{aligned} PCR_k &= \min\left[\frac{USL - \bar{\bar{x}}}{3\hat{\sigma}}, \frac{\bar{\bar{x}} - LSL}{3\hat{\sigma}}\right] \\ &= \min\left[\frac{40 - 33.21}{3(2.15)} = 1.06, \frac{33.21 - 20}{3(2.15)} = 2.04\right] = 1.05 \end{aligned}$$

Note that if $PCR = PCR_k$, the process is centered at the nominal dimension. Because $PCR_k = 1.05$ for the vane-opening process and $PCR = 1.55$, the process is obviously running off-center, as was first noted in Figures 15.10 and 15.13. This off-center operation was ultimately traced to an oversized wax tool. Changing the tooling resulted in a substantial improvement in the process.

The fractions of nonconforming output (or fallout) below the lower specification limit and above the upper specification limit are often of interest. Suppose that the output from a normally distributed process in statistical control is denoted as X. The fractions are determined from

$$P(X < LSL) = P\left(Z < \frac{LSL - \mu}{\sigma}\right) \qquad P(X > USL) = P\left(Z > \frac{USL - \mu}{\sigma}\right)$$

EXAMPLE 15.3 | Electrical Current

For an electronic manufacturing process a current has specifications of 100 ± 10 milliamperes. The process mean μ and standard deviation σ are 107.0 and 1.5, respectively. The process mean is nearer to the *USL*. Consequently,

$$PCR = \frac{110 - 90}{6(1.5)} = 2.22$$

and

$$PCR_k = \frac{110 - 107}{3(1.5)} = 0.67$$

The small PCR_k indicates that the process is likely to produce currents outside the specification limits. From the normal distribution in Appendix Table II,

$$P(X < LSL) = P\left(Z < \frac{90 - 107}{1.5}\right) = P(Z < -11.33) \cong 0$$

$$P(X > USL) = P\left(Z > \frac{110 - 107}{1.5}\right) = P(Z > 2) = 0.023$$

Practical Interpretation: The probability a current is less than the *LSL* is nearly zero. However, the relatively large probability of exceeding the *USL* is a warning of potential problems with this criterion even if none of the measured observations in a preliminary sample exceeds this limit. The PCR_k would improve if the process mean were centered in the specifications at 100 milliamperes.

TABLE 15.4 PCR Related to PPM for a Normally Distributed Process

	PPM	
PCR	Mean Centered	Mean Shifted 1.5σ
0.5	133,614.4	501,349.9
0.67	44,431.2	305,249.8
0.75	24,448.9	226,715.8
1	2,699.8	66,810.6
1.25	176.8	12,224.5
1.33	66.1	6,387.2
1.5	6.8	1,349.9
1.67	0.5	224.1
1.75	0.2	88.4
2	0.0	3.4

We emphasize that the fraction-nonconforming calculation assumes that the observations are normally distributed and the process is in control. Departures from normality can seriously affect the results. The calculation should be interpreted as an approximate guideline for process performance. To make matters worse, μ and σ need to be estimated from the data available, and a small sample size can result in poor estimates that further degrade the calculation.

Table 15.4 relates fallout in **parts per million (PPM)** for a normally distributed process in statistical control to the value of PCR. The table shows PPM for a centered process and for one with a 1.5σ shift in the process mean. Many U.S. companies use $PCR = 1.33$ as a minimum acceptable target and $PCR = 1.66$ as a minimum target for strength, safety, or critical characteristics.

Some companies require that internal processes and those at suppliers achieve a $PCR_k = 2.0$. Figure 15.15 illustrates a process with $PCR = PCR_k = 2.0$. Assuming a normal distribution, the calculated fallout for this process is 0.0018 parts per million. A process with $PCR_k = 2.0$ is referred to as a **6-sigma process** because the distance from the process mean to the nearest specification is 6 standard deviations. The reason that such a large process capability is often required is that maintaining a process mean at the center of the specifications for long periods of time is difficult. A common model that is used to justify the importance of a 6-sigma process is illustrated by referring to Figure 15.15. If the process mean shifts off-center by 1.5 standard deviations, the PCR_k decreases to

$$PCR_k = \frac{USL - \mu}{3\sigma} = \frac{6\sigma - 1.5\sigma}{3\sigma} = \frac{4.5\sigma}{3\sigma} = 1.5$$

Assuming a normally distributed process, the fallout of the shifted process is **3.4 parts per million**. Consequently, the mean of a 6-sigma process can shift 1.5 standard deviations from the center of the specifications and still maintain a minimal fallout.

In addition, some U.S. companies, particularly the automobile industry, have adopted the terminology $C_p = PCR$ and $C_{pk} = PCR_k$. Because C_p has another meaning in statistics (in multiple regression), we prefer the traditional notation PCR and PCR_k.

We repeat that process capability calculations are meaningful only for stable processes; that is, processes that are in control. A process capability ratio indicates whether or not the natural or chance variability in a process is acceptable relative to the specifications.

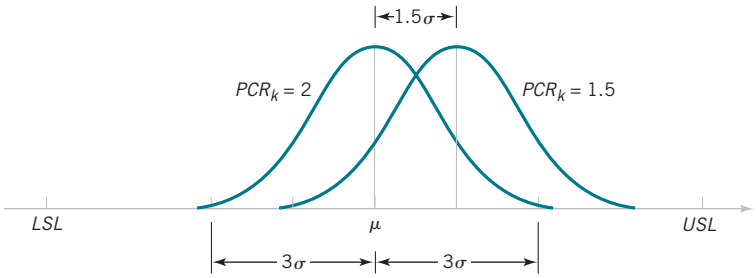


FIGURE 15.15

Mean of a 6-sigma process shifts by 1.5 standard deviations.

15.6 Attribute Control Charts

15.6.1 *P* Chart (Control Chart for Proportions)

Often it is desirable to classify a product as either defective or nondefective on the basis of comparison with a standard. For example, the diameter of a ball bearing may be checked by determining whether it passes through a gauge consisting of circular holes cut in a template. This kind of measurement would be much simpler than directly measuring the diameter with a device such as a micrometer. Control charts for attributes are used in these situations. Attribute control charts often require a considerably larger sample size than do their variable measurements counterparts. In this section, we discuss the **fraction-defective control chart**, or ***P* chart**. Sometimes the P chart is called the *control chart for fraction nonconforming*.

At each sample time, a random sample of n units is selected. Suppose that D is the number of defective units in the sample. We assume that D is a binomial random variable with unknown parameter p. The fraction defective

$$P = \frac{D}{n}$$

of each sample is plotted on the chart. Furthermore, from the binomial distribution, the variance of the statistic P is

$$\sigma_P^2 = \frac{p(1-p)}{n}$$

Therefore, a P chart for fraction defective could be constructed using p as the center line and control limits at

$$UCL = p + 3\sqrt{\frac{p(1-p)}{n}} \qquad LCL = p - 3\sqrt{\frac{p(1-p)}{n}} \tag{15.21}$$

However, the true process fraction defective is almost always unknown and must be estimated using the data from preliminary samples.

Suppose that m preliminary samples each of size n are available, and let D_i be the number of defectives in the ith sample. Then $P_i = D_i/n$ is the sample fraction defective in the ith sample. The average fraction defective is

$$\overline{P} = \frac{1}{m}\sum_{i=1}^{m} P_i = \frac{1}{mn}\sum_{i=1}^{m} D_i \tag{15.22}$$

Now $\overline{P}$ may be used as an estimator of p in the center line and control limit formulas.

P Chart

The center line and upper and lower control limits for the *P* chart are

$$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \qquad CL = \bar{p} \qquad LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \tag{15.23}$$

where $\bar{p}$ is the observed value of the average fraction defective.

These control limits are based on the normal approximation to the binomial distribution. When p is small, the normal approximation may not always be adequate. In such cases, we may use control limits obtained directly from a table of binomial probabilities. If $\bar{p}$ is small, the lower control limit obtained from the normal approximation may be a negative number. If this should occur, it is customary to consider zero as the lower control limit.

EXAMPLE 15.4 | Ceramic Substrate

We wish to construct a fraction-defective control chart for a ceramic substrate production line. We have 20 preliminary samples, each of size 100; the number of defectives in each sample is shown in Table 15.5. Assume that the samples are numbered in the sequence of production. Note that $\bar{p} = (800/2000) = 0.40$; therefore, the trial parameters for the control chart are

$$UCL = 0.40 + 3\sqrt{\frac{(0.40)(0.60)}{100}} = 0.55 \qquad CL = 0.40$$

$$LCL = 0.40 - 3\sqrt{\frac{(0.40)(0.60)}{100}} = 0.25$$

The control chart is shown in Figure 15.16. All samples are in control. If they were not, we would search for assignable causes of variation and revise the limits accordingly. This chart can be used for controlling future production.

Practical Interpretation: Although this process exhibits statistical control, its defective rate ($\bar{p} = 0.40$) is very poor. We should take appropriate steps to investigate the process to determine why such a large number of defective units is being produced. Defective units should be analyzed to determine the specific types of defects present. Once the defect types are known, process changes should be investigated to determine their impact on defect levels. Designed experiments may be useful in this regard.

TABLE 15.5 Number of Defectives in Samples of 100 Ceramic Substrates

Sample	No. of Defectives	Sample	No. of Defectives
1	44	11	36
2	48	12	52
3	32	13	35
4	50	14	41
5	29	15	42
6	31	16	30
7	46	17	46
8	52	18	38
9	44	19	26
10	48	20	30

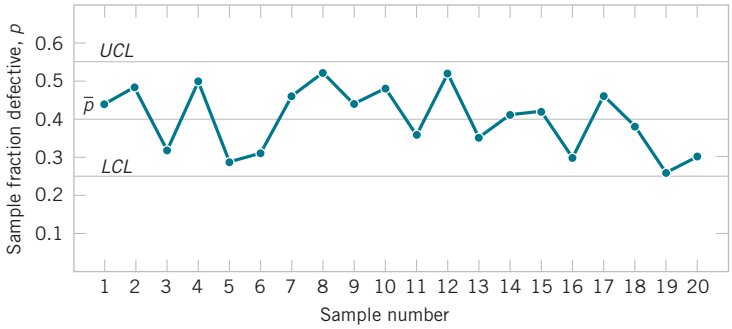


FIGURE 15.16

***P* chart for a ceramic substrate.**

Computer software also produces an ***NP* chart**. This is just a control chart of $nP = D$, the number of defectives in a sample. The points, center line, and control limits for this chart are simply multiples (times n) of the corresponding elements of a P chart. The use of an NP chart avoids the fractions in a P chart, but it is otherwise equivalent.

15.6.2 *U* Chart (Control Chart for Defects per Unit)

It is sometimes necessary to monitor the number of defects in a unit of product rather than the fraction defective. For example, a hospital might record the number of cases of infection per month, or a semiconductor manufacturer might record the number of large contamination particles per wafer. In these situations, we may use the **defects-per-unit chart** or ***U* chart**. If each subgroup consists of n units and there are C total defects in the subgroup, then,

$$U = \frac{C}{n}$$

is the average number of defects per unit. A U chart may be constructed for such data.

Many defects-per-unit situations can be modeled by the Poisson distribution. Suppose that the number of defects in a unit is a Poisson random variable with mean λ. The variance also equals λ. Each point on the chart is an observed value of U, the average number of defects per unit from a sample of n units. The mean of U is λ, and the variance of U is λ/n. Therefore, the control limits for the U chart with known λ are:

$$UCL = \lambda + 3\sqrt{\frac{\lambda}{n}} \qquad LCL = \lambda - 3\sqrt{\frac{\lambda}{n}} \tag{15.24}$$

Suppose that m preliminary samples each of size n are available, and let C_i be the number of defectives in the ith sample. Then $U_i = C_i/n$ is the average number of defects per unit in the ith sample. The estimator of the average number of defects per unit is

$$\overline{U} = \frac{1}{m}\sum_{i=1}^{m} U_i = \frac{1}{mn}\sum_{i=1}^{m} C_i \tag{15.25}$$

Now $\overline{U}$ is used as an estimator of λ in the center line and control limit formulas.

U Chart

The center line and upper and lower control limits on the U chart are

$$UCL = \overline{u} + 3\sqrt{\frac{\overline{u}}{n}} \qquad CL = \overline{u} \qquad LCL = \overline{u} - 3\sqrt{\frac{\overline{u}}{n}} \tag{15.26}$$

where $\overline{u}$ is the average number of defects per unit.

These control limits are based on the normal approximation to the Poisson distribution. When λ is small, the normal approximation may not always be adequate. In such cases, we may use control limits obtained directly from a table of Poisson probabilities. If $\overline{u}$ is small, the lower control limit obtained from the normal approximation may be a negative number. If this should occur, it is customary to use zero as the lower control limit.

EXAMPLE 15.5 | Printed Circuit Boards

Printed circuit boards are assembled by a combination of manual assembly and automation. Surface mount technology (SMT) is used to make the mechanical and electrical connections of the components to the board. Every hour, five boards are selected and inspected for process-control purposes. The number of defects in each sample of five boards is noted. Results for 20 samples are shown in Table 15.6.

The center line for the U chart is

$$\bar{u} = \frac{1}{20}\sum_{i=1}^{20} u_i = \frac{32.0}{20} = 1.6$$

and the upper and lower control limits are

$$UCL = \bar{u} + 3\sqrt{\frac{\bar{u}}{n}} = 1.6 + 3\sqrt{\frac{1.6}{5}} = 3.3$$

$$LCL = \bar{u} - 3\sqrt{\frac{\bar{u}}{n}} = 1.6 - 3\sqrt{\frac{1.6}{5}} < 0$$

The control chart is plotted in Figure 15.17. Because LCL is negative, it is set to 0. From the control chart in Figure 15.17, we see that the process is in control.

Practical Interpretation: Eight defects per group of five circuit boards are too many (about 8/5 = 1.6 defects/board), and the process needs improvement. An investigation of the specific types of defects found on the printed circuit board needs to be made. This usually suggests potential avenues for process improvement.

TABLE 15.6 Number of Defects in Samples of Five Printed Circuit Boards

Sample	Number of Defects	Defects per Unit u_i	Sample	Number of Defects	Defects per Unit u_i
1	6	1.2	11	9	1.8
2	4	0.8	12	15	3.0
3	8	1.6	13	8	1.6
4	10	2.0	14	10	2.0
5	9	1.8	15	8	1.6
6	12	2.4	16	2	0.4
7	16	3.2	17	7	1.4
8	2	0.4	18	1	0.2
9	3	0.6	19	7	1.4
10	10	2.0	20	13	2.6

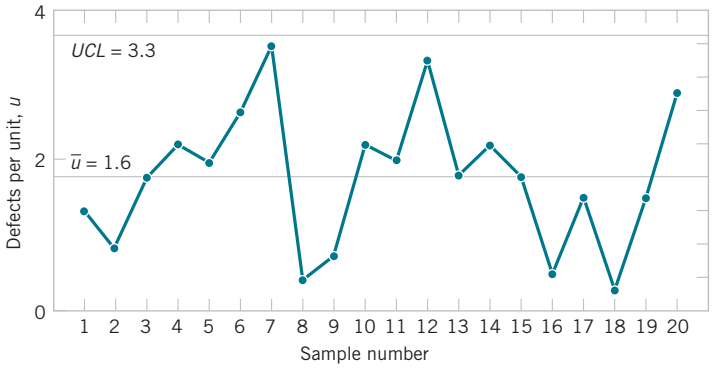


FIGURE 15.17

***U* chart of defects per unit on printed circuit boards.**

Computer software also produces a ***C* chart**. This is just a control chart of C, the total of defects in a sample. The points, center line, and control limits for this chart are simply multiples (times n) of the corresponding elements of a U chart. The use of a C chart avoids the fractions that can occur in a U chart, but it is otherwise equivalent.

15.7 Control Chart Performance

Specifying the control limits is one of the critical decisions that must be made in designing a control chart. By moving the control limits farther from the center line, we decrease the risk of a type I error—that is, the risk of a point falling beyond the control limits, indicating an out-of-control condition when no assignable cause is present. However, widening the control limits also increases the risk of a type II error—that is, the risk of a point falling between the control limits when the process is really out of control. If we move the control limits closer to the center line, the opposite effect is obtained: The risk of type I error increases, and the risk of type II error decreases.

The control limits on a Shewhart control chart are customarily located a distance of plus or minus 3 standard deviations of the statistic plotted on the chart from the center line. That is, the constant k in Equation 15.1 should be set equal to 3. These limits are called **3-sigma control limits**.

A way to evaluate decisions regarding sample size and sampling frequency is through the **average run length (ARL)** of the control chart. Essentially, the ARL is the average number of points plotted to signal an out-of-control condition. For any Shewhart control chart, the ARL can be calculated from the mean of a geometric random variable. Suppose that p is the probability that any point exceeds the control limits. Then

Average Run Length

$$\text{ARL} = \frac{1}{p} \tag{15.27}$$

Thus, for an $\overline{X}$ chart with 3-sigma limits, $p = 0.0027$ is the probability that a normally distributed point falls outside the limits when the process is in control, so

$$\text{ARL} = \frac{1}{p} = \frac{1}{0.0027} \cong 370$$

is the average run length of the $\overline{X}$ chart when the process is in control. That is, even if the process remains in control, an out-of-control signal is generated every 370 points on average.

Consider the piston ring process discussed in Section 15.2.2, and suppose that we are sampling every hour. Thus, we have a **false alarm** about every 370 hours on average. Suppose that we are using a sample size of $n = 5$ and that when the process goes out of control, the mean shifts to 74.0135 millimeters. Then, the probability that $\overline{X}$ falls between the control limits of Figure 15.3 is equal to

$$\begin{aligned} &P\,[73.9865 \le \overline{X} \le 74.0135 \text{ when } \mu = 74.0135] \\ &= P\left[\frac{73.9865 - 74.0135}{0.0045} \le Z \le \frac{74.0135 - 74.0135}{0.0045}\right] \\ &= P[-6 \le Z \le 0] = 0.5 \end{aligned}$$

Therefore, p in Equation 15.27 is 0.50, and the out-of-control ARL is

$$\text{ARL} = \frac{1}{p} = \frac{1}{0.5} = 2$$

TABLE 15.7 **Average Run Length (ARL) for an $\overline{X}$ Chart with 3-Sigma Control Limits**

Magnitude of Process Shift	ARL $n = 1$	ARL $n = 4$
0	370.4	370.4
0.5σ	155.2	43.9
1.0σ	43.9	6.3
1.5σ	15.0	2.0
2.0σ	6.3	1.2
3.0σ	2.0	1.0

That is, the control chart requires two samples to detect the process shift, on the average, so 2 hours is expected to elapse between the shift and its detection (*again, on the average*). Suppose that this approach is unacceptable because production of piston rings with a mean diameter of 74.0135 millimeters results in excessive scrap costs and delays final engine assembly. How can we reduce the time needed to detect the out-of-control condition? One method is to sample more frequently. For example, if we sample every half hour, only one hour elapses (on the average) between the shift and its detection. The second possibility is to increase the sample size. For example, if we use $n = 10$, the control limits in Figure 15.3 narrow to 73.9905 and 74.0095. The probability of $\overline{X}$ falling between the control limits when the process mean is 74.0135 millimeters is approximately 0.1, so $p = 0.9$, and the out-of-control ARL is

$$\text{ARL} = \frac{1}{p} = \frac{1}{0.9} = 1.11$$

Thus, the larger sample size would allow the shift to be detected about twice as quickly as the smaller one. If it became important to detect the shift in approximately the first hour after it occurred, two control chart designs would work:

Design 1	Design 2
Sample size: $n = 5$	Sample size: $n = 10$
Sampling frequency: every half hour	Sampling frequency: every hour

Table 15.7 provides average run lengths for an $\overline{X}$ chart with 3-sigma control limits. The average run lengths are calculated for shifts in the process mean from 0 to 3.0σ and for sample sizes of $n = 1$ and $n = 4$ by using $1/p$, where p is the probability that a point plots outside of the control limits (based on a normal distribution). Figure 15.18 illustrates a shift in the process mean of 2σ.

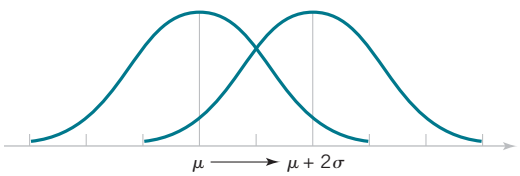


FIGURE 15.18
Process mean shift of 2σ.

15.8 Time-Weighted Charts

In Sections 15.3 and 15.4 we presented basic types of Shewhart control charts. A major disadvantage of any Shewhart control chart is that it is relatively insensitive to small shifts in the process, say, on the order of about 1.5σ or less. One reason is that a Shewhart chart uses only the information in the last plotted point and ignores the information in the sequence of points. This problem can be addressed to some extent by adding criteria such as the Western Electric rules to a Shewhart chart, but the use of these rules reduces the simplicity and ease of interpretation of the chart. These rules would also cause the in-control average run length of a Shewhart chart to drop below 370. This increase in the false alarm rate can have serious practical consequences.

An effective alternative to the Shewhart control chart is a **time-weighted chart**, which integrates data over several time periods. Such a chart has much better performance (in terms of ARL) to detect small shifts than the Shewhart chart, but it does not cause the in-control ARL to drop substantially.

15.8.1 Exponentially Weighted Moving-Average Control Chart

Data collected in time order are often averaged over several time periods. For example, economic data are often presented as an average over the last four quarters. That is, at time t, the average of the last four measurements can be written as

$$\bar{x}_t = \frac{1}{4}x_t + \frac{1}{4}x_{t-1} + \frac{1}{4}x_{t-2} + \frac{1}{4}x_{t-3}$$

This average places weight of 1/4 on each of the most recent observations and zero weight on older observations. It is called a **moving average** and in this case, a *window* of size 4 is used. An average of the recent data is used to smooth the noise in the data to generate a better estimate of the process mean than only the most recent observation.

For statistical process control, rather than use a fixed window size, it is useful to place the most weight on the most recent observation or subgroup average and then gradually decrease the weights on older observations. An average of this type can be constructed by a multiplicative decrease in the weights. Let $0 < \lambda \leq 1$ denote a constant and μ_0 denote the process target or historical mean. Suppose that samples of size $n \geq 1$ are collected and $\bar{x}_t$ is the average of the sample at time t. The **exponentially weighted moving-average (EWMA)** is

$$\begin{aligned} z_t &= \lambda\bar{x}_t + \lambda(1-\lambda)\bar{x}_{t-1} + \lambda(1-\lambda)^2\bar{x}_{t-2} + \cdots + \lambda(1-\lambda)^{t-1}\bar{x}_1 + (1-\lambda)^t\mu_0 \\ &= \sum_{k=0}^{t-1}\lambda(1-\lambda)^k\bar{x}_{t-k} + (1-\lambda)^t\mu_0 \end{aligned}$$

Each older observation has its weight decreased by the factor $(1-\lambda)$. The weight on the starting value μ_0 is selected so that the weights sum to 1. An EWMA is also sometimes called a **geometric average**.

The value of λ determines the compromise between noise reduction and inertia to a change. For example, the series of weights when $\lambda = 0.8$ are

$$0.8, 0.16, 0.032, 0.0064, 0.00128, \ldots$$

and when $\lambda = 0.2$, the weights are

$$0.2, 0.16, 0.128, 0.1024, 0.0819, \ldots$$

When $\lambda = 0.8$, the weights decrease rapidly. Most of the weight is placed on the most recent observation with modest contributions to the EWMA from older measurements. In this case, the EWMA does not average noise much (so control limits are wider), but it has small inertia. The result is a chart that is more sensitive to a larger change (similar to a Shewhart chart). However,

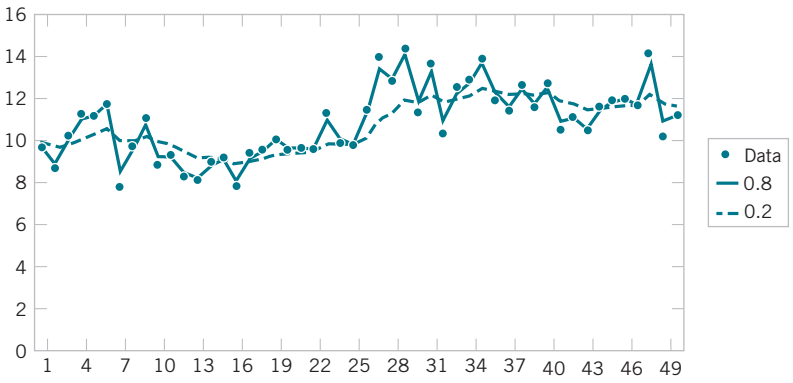


FIGURE 15.19

EWMAs with $\lambda = 0.8$ and $\lambda = 0.2$ show a compromise between a smoother curve and inertia to a shift.

when $\lambda = 0.2$, the weights decrease much more slowly and the EWMA has substantial contributions from the more recent observations. In this case, the EWMA averages noise more (so control limits are narrower), but it has greater inertia. The result is a chart that is more sensitive to a smaller change. Figure 15.19 displays a series of observations with a mean shift in the middle on the series. Notice that the EWMA with $\lambda = 0.2$ smooths the data more but that the EWMA with $\lambda = 0.8$ adjusts the estimate to the mean shift more quickly.

It appears that it is difficult to calculate an EWMA because at every time t a new weighted average of all previous data is required. However, there is an easy method to calculate z_t based on a simple recursive equation. Let $z_0 = \mu_0$. Then it can be shown that

EWMA Update Equation

$$z_t = \lambda \bar{x}_t + (1 - \lambda) z_{t-1} \tag{15.28}$$

Consequently, only a brief computation is needed at each time t.

To develop a control chart from an EWMA, control limits are needed for Z_t. The control limits are defined in a straightforward manner. They are placed at 3 standard deviations around the mean of the plotted statistic Z_t. This follows the general approach for a control chart in Equation 15.1. An EWMA control chart may be applied to individual measurements or to subgroup averages. Formulas here are developed for the more general case with an average from a subgroup of size n. For individual measurements, $n = 1$.

Because Z_t is a linear function of the independent observations $X_1, X_2, \ldots, X_t$ (and μ_0), the results from Chapter 5 can be used to show that

$$E(Z_t) = \mu_0 \quad \text{and} \quad V(Z_t) = \frac{\sigma^2}{n} \frac{\lambda}{2 - \lambda} \left[1 - (1 - \lambda)^{2t}\right]$$

where n is the subgroup size. Therefore, an EWMA control chart uses estimates of μ_0 and σ in the following formulas:

EWMA Control Chart

$$\begin{aligned} LCL &= \mu_0 - 3\frac{\sigma}{\sqrt{n}} \sqrt{\frac{\lambda}{2 - \lambda} \left[1 - (1 - \lambda)^{2t}\right]} \\ CL &= \mu_0 \\ UCL &= \mu_0 + 3\frac{\sigma}{\sqrt{n}} \sqrt{\frac{\lambda}{2 - \lambda} \left[1 - (1 - \lambda)^{2t}\right]} \end{aligned} \tag{15.29}$$

Note that the control limits are not of equal width about the center line. The control limits are calculated from the variance of Z_t and that changes with time. However, for large t, the variance of Z_t converges to

$$\lim_{t\to\infty} V(Z_t) = \frac{\sigma^2}{n}\left(\frac{\lambda}{2-\lambda}\right)$$

so that the control limits tend to be parallel lines about the center line as t increases.

The parameters μ_0 and σ are estimated by the same statistics used in $\overline{X}$ or X charts. That is, for subgroups

$$\hat{\mu}_0 = \overline{\overline{X}} \quad \text{and} \quad \hat{\sigma} = \overline{R}/d_2 \quad \text{or} \quad \hat{\sigma} = \overline{S}/c_4$$

and for $n = 1$

$$\hat{\mu}_0 = \overline{X} \quad \text{and} \quad \hat{\sigma} = \overline{MR}/1.128$$

EXAMPLE 15.6 | Chemical Process Concentration EWMA

Consider the concentration data shown in Table 15.3. Construct an EWMA control chart with $\lambda = 0.2$ with $n = 1$. It was determined that $\overline{x} = 99.1$ and $\overline{mr} = 2.59$. Therefore, $\hat{\mu}_0 = 99.1$ and $\hat{\sigma} = 2.59/1.128 = 2.30$. The control limits for z_1 are

$$LCL = 99.1 - 3(2.30)\sqrt{\frac{0.2}{2-0.2}\left[1-(1-0.2)^2\right]} = 98.19$$

$$UCL = 99.1 + 3(2.30)\sqrt{\frac{0.2}{2-0.2}\left[1-(1-0.2)^2\right]} = 100.01$$

The first few values of z_t along with the corresponding control limits are

t	1	2	3	4	5
x_t	102	94.8	98.3	98.4	102
z_t	99.68	98.7	98.62	98.58	99.26
LCL	97.72	97.33	97.12	97	96.93
UCL	100.48	100.87	101.08	101.2	101.27

The chart generated by computer software is shown in Figure 15.20. Notice that the control limits widen as time increases but quickly stabilize. Each point is within its set of corresponding control limits so there are no signals from the chart.

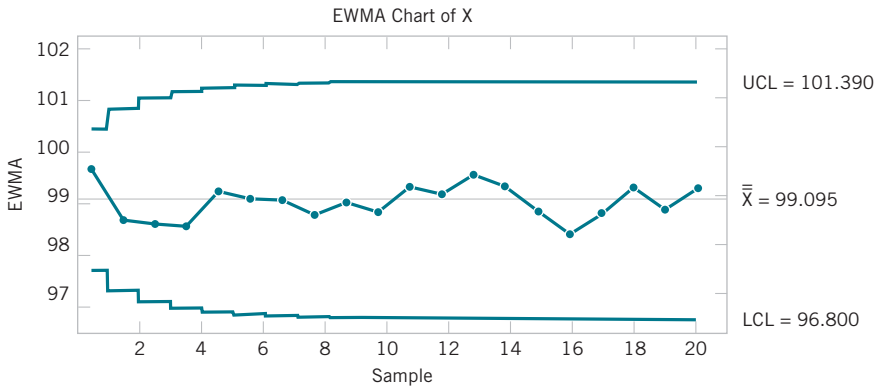


FIGURE 15.20

EWMA control chart for the chemical process concentration data from computer software.

The points plotted on an EWMA control chart are not independent. Therefore, run rules should not be applied to an EWMA control chart. Information in the history of the data that is considered by run rules is to a large extent incorporated into the EWMA that is calculated at each time t.

TABLE 15.8 Average Run Lengths for an EWMA Control Chart

Shift in Mean (multiple of $\sigma_{\bar{X}}$)	$\lambda = 0.5$ $L = 3.07$	$\lambda = 0.1$ $L = 2.81$
0	500	500
0.25	255	106
0.5	88.8	31.3
0.75	35.9	15.9
1	17.5	10.3
1.5	6.53	6.09
2	3.63	4.36
3	1.93	2.87

The value of λ is usually chosen from the range $0.1 < \lambda < 0.5$. A common choice is $\lambda = 0.2$. Smaller values for λ provide more sensitivity for small shifts and larger values better tune the chart for larger shifts. This performance can be seen in the ARLs in Table 15.8. These calculations are more difficult than those used for Shewhart charts, and details are omitted. Here, $\lambda = 0.1$ and 0.5 are compared. The multiplier of the standard deviation, denoted L in the table, is adjusted so that the ARL equals 500 for both choices for λ. That is, the control limits are placed at $E(Z_t) \pm L\sqrt{V(Z_t)}$, and L is selected such that the ARL without a mean shift is 500 in both cases.

The EWMA ARLs in the table indicate that a smaller value for λ is preferred when the magnitude of the shift is small. Also, the EWMA performance is in general much better than results for a Shewhart control chart (in Table 15.7). However, these are average results. At the time of an increase in the process mean, z_t might be negative, and there would be some performance penalty to first increase z_t to near zero and then further increase it to a signal above the *UCL*. A more refined analysis can be used to quantify this penalty, but the conclusion is that the EWMA penalty is moderate to small in most applications.

15.8.2 Cumulative Sum Control Chart

Another alternative to the Shewhart control chart is the **cumulative sum control chart (CUSUM)**. This section illustrates the use of the CUSUM for sample averages and individual measurements. CUSUM charts for other sample statistics are also available.

The CUSUM chart plots the cumulative sums of the deviations of the sample values from a target value. For example, suppose that samples of size $n \geq 1$ are collected, and $\bar{X}_j$ is the average of the jth sample. Then if μ_0 is the target for the process mean, the cumulative sum control chart is formed by plotting the quantity

$$S_i = \sum_{j=1}^{i} (\bar{X}_j - \mu_0) \tag{15.30}$$

against the sample number i. Now S_i is called the cumulative sum up to and including the ith sample. Because they combine information from *several* samples, cumulative sum charts are more effective than Shewhart charts for detecting small process shifts. Furthermore, they are particularly effective with samples of $n = 1$. This makes the cumulative sum control chart a good

candidate for use in the chemical and process industries in which rational subgroups are frequently of size 1, as well as in discrete parts manufacturing with automatic measurement of each part and online control using a computer directly at the work center.

If the process remains in control at the target value μ_0, the cumulative sum defined in Equation 15.28 should fluctuate around zero. However, if the mean shifts upward to some value $\mu_1 > \mu_0$, for example, an upward or positive drift develops in the cumulative sum S_i. Conversely, if the mean shifts downward to some $\mu_1 < \mu_0$, a downward or negative drift in S_i develops. Therefore, if a trend develops in the plotted points either upward or downward, we should consider this as evidence that the process mean has shifted, and a search for the assignable cause should be performed.

This theory can easily be demonstrated by applying the CUSUM to the chemical process concentration data in Table 15.3. Because the concentration readings are individual measurements, we would take $\overline{X}_j = X_j$ in computing the CUSUM. Suppose that the target value for the concentration is $\mu_0 = 99$. Then the CUSUM is

$$\begin{aligned} S_i &= \sum_{j=1}^{i}(X_j - 99) = (X_i - 99) + \sum_{j=1}^{i-1}(X_j - 99) \\ &= (X_i - 99) + S_{i-1} \end{aligned}$$

Table 15.9 shows the computing values, s_i's for this CUSUM, for which the starting value of the CUSUM, s_0, is taken to be zero. Figure 15.21 plots the CUSUM from the last column of Table 15.9. Notice that the CUSUM fluctuates around the value of 0.

TABLE 15.9 **CUSUM Computations for the Chemical Process Concentration Data in Table 15-3**

Observation, i	x_i	$x_i - 99$	$s_i = (x_i - 99) + s_{i-1}$
1	102.0	3.0	3.0
2	94.8	−4.2	−1.2
3	98.3	−0.7	−1.9
4	98.4	−0.6	−2.5
5	102.0	3.0	0.5
6	98.5	−0.5	0.0
7	99.0	0.0	0.0
8	97.7	−1.3	−1.3
9	100.0	1.0	−0.3
10	98.1	−0.9	−1.2
11	101.3	2.3	1.1
12	98.7	−0.3	0.8
13	101.1	2.1	2.9
14	98.4	−0.6	2.3
15	97.0	−2.0	0.3
16	96.7	−2.3	−2.0
17	100.3	1.3	−0.7
18	101.4	2.4	1.7
19	97.2	−1.8	−0.1
20	101.0	2.0	1.9

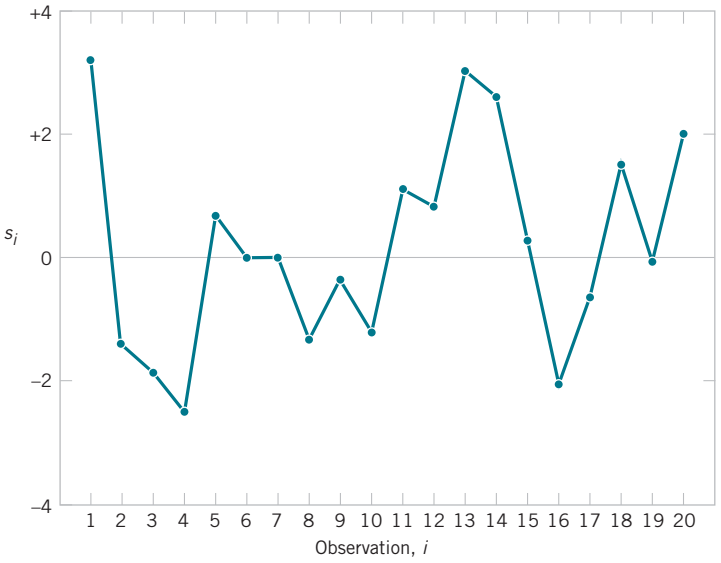


FIGURE 15.21

Plot of the cumulative sums for the concentration data.

The graph in Figure 15.21 is not a control chart because it lacks control limits. There are two general approaches to devising control limits for CUSUMs. The older of these two methods is the V-mask procedure. A typical V mask is shown in Figure 15.22(a). It is a V-shaped notch in a plane that can be placed at different locations on the CUSUM chart. The decision procedure consists of placing the V mask on the cumulative sum control chart with the point O on the last value of s_i and the line OP parallel to the horizontal axis. If all the previous cumulative sums, s_1, s_2, … , s_{i-1}, lie within the two arms of the V mask, the process is in control. The arms are the lines that make angles θ with segment OP in Figure 15.22(a) and are assumed to extend infinitely in length. However, if any s_i lies outside the arms of the mask, the process is considered to be out of control. In actual use, the V mask would be applied to each new point on the CUSUM chart

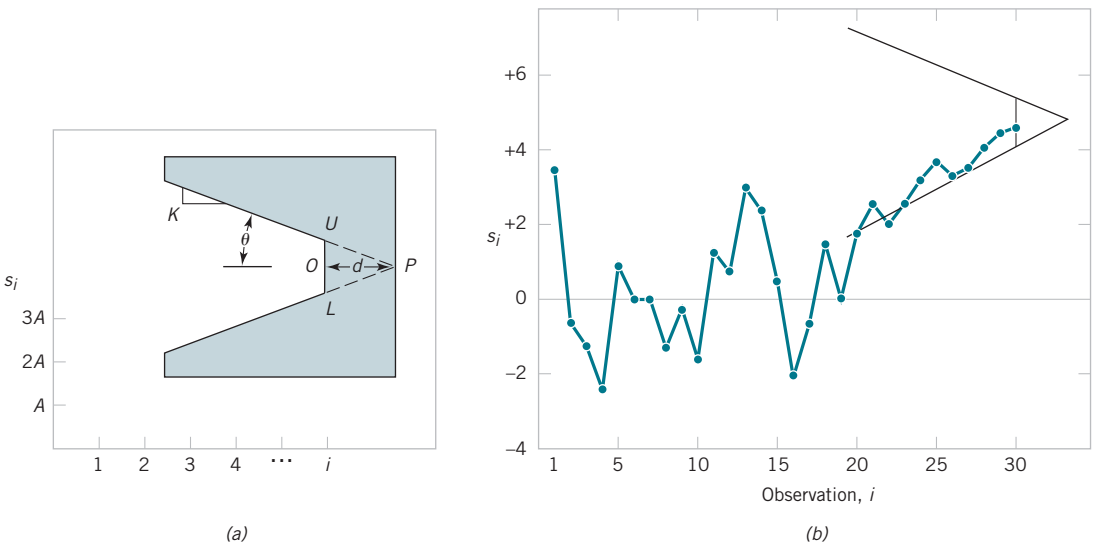


FIGURE 15.22

The cumulative sum control chart. (a) The V-mask and scaling. (b) The cumulative sum control chart in operation.

as soon as it was plotted. The example in Figure 15.22(b) indicates an upward shift in the mean because at least one of the points that occurred earlier than sample 22 now lies below the lower arm of the mask when the V mask is centered on sample 30. If the point lies above the upper arm, a downward shift in the mean is indicated. Thus, the V mask forms a visual frame of reference similar to the control limits on an ordinary Shewhart control chart. For the technical details of designing the V mask, see Montgomery (2013).

Although some computer programs plot CUSUMs with the V-mask control scheme, we believe that the other approach to CUSUM control, the **tabular CUSUM**, is superior. The tabular procedure is particularly attractive when the CUSUM is implemented on a computer.

Let $S_H(i)$ be an upper one-sided CUSUM for period i and $S_L(i)$ be a lower one-sided CUSUM for period i. These quantities are calculated from

CUSUM Control Chart

$$s_H(i) = \max[0, \bar{x}_i - (\mu_0 + K) + s_H(i-1)] \tag{15.31}$$

and

$$s_L(i) = \max[0, (\mu_0 - K) - \bar{x}_i + s_L(i-1)] \tag{15.32}$$

where the starting values $s_H(0) = s_L(0) = 0$.

In Equations 15.31 and 15.32, K is called the **reference value**, which is usually chosen about halfway between the target μ_0 and the value of the mean corresponding to the out-of-control state, $\mu_1 = \mu_0 + \Delta$. That is, K is about one-half the magnitude of the shift we are interested in, or

$$K = \frac{\Delta}{2}$$

Notice that $S_H(i)$ and $S_L(i)$ accumulate deviations from the target value that are greater than K, with both quantities reset to zero upon becoming negative. If either $S_H(i)$ or $S_L(i)$ exceeds a constant H, the process is out of control. This constant H is usually called the **decision interval**.

EXAMPLE 15.7 | Chemical Process Concentration Tabular CUSUM

We illustrate the tabular CUSUM with the chemical process concentration data in Table 15.9. The process target is $\mu_0 = 99$, and we use $K = 1$ as the reference value and $H = 10$ as the decision interval. The reasons for these choices are explained later.

Table 15.10 shows the tabular CUSUM scheme for the chemical process concentration data. To illustrate the calculations, note that

$$\begin{aligned} S_H(i) &= \max[0, x_i - (\mu_0 + K) + S_H(i-1)] \\ &= \max[0, x_i - (99 + 1) + S_H(i-1)] \\ &= \max[0, x_i - (100) + S_H(i-1)] \end{aligned}$$

$$\begin{aligned} S_L(i) &= \max[0, (\mu_0 - K) - x_i + S_L(i-1)] \\ &= \max[0, (99 - 1) - x_i + S_L(i-1)] \\ &= \max[0, 98 - x_i + S_L(i-1)] \end{aligned}$$

Therefore, for observation 1, the CUSUMs are

$$\begin{aligned} s_H(1) &= \max[0, x_1 - 100 + s_H(0)] \\ &= \max[0, 102.0 - 100 + 0] = 2.0 \end{aligned}$$

and

$$\begin{aligned} s_L(1) &= \max[0, 98 - x_1 + s_L(0)] \\ &= \max[0, 98 - 102.0 + 0] = 0 \end{aligned}$$

as shown in Table 15.10. The quantities n_H and n_L in Table 15.10 indicate the number of periods in which the CUSUM $s_H(i)$ or $s_L(i)$ have been nonzero. Notice that the CUSUMs in this example never exceed the decision interval $H = 10$. We would therefore conclude that the process is in control.

Next Steps: The limits for the CUSUM charts may be used to continue to operate the chart in order to monitor future productions.

TABLE 15.10 Tabular CUSUM for the Chemical Process Concentration Data

Observation i	Upper CUSUM				Lower CUSUM		
	x_i	$x_i - 100$	$s_H(i)$	n_H	$98 - x_i$	$s_L(i)$	n_L
1	102.0	2.0	2.0	1	−4.0	0.0	0
2	94.8	−5.2	0.0	0	3.2	3.2	1
3	98.3	−1.7	0.0	0	−0.3	2.9	2
4	98.4	−1.6	0.0	0	−0.4	2.5	3
5	102.0	2.0	2.0	1	−4.0	0.0	0
6	98.5	−1.5	0.5	2	−0.5	0.0	0
7	99.0	−1.0	0.0	0	−1.0	0.0	0
8	97.7	−2.3	0.0	0	0.3	0.3	1
9	100.0	0.0	0.0	0	−2.0	0.0	0
10	98.1	−1.9	0.0	0	−0.1	0.0	0
11	101.3	1.3	1.3	1	−3.3	0.0	0
12	98.7	−1.3	0.0	0	−0.7	0.0	0
13	101.1	1.1	1.1	1	−3.1	0.0	0
14	98.4	−1.6	0.0	0	−0.4	0.0	0
15	97.0	−3.0	0.0	0	1.0	1.0	1
16	96.7	−3.3	0.0	0	1.3	2.3	2
17	100.3	0.3	0.3	1	−2.3	0.0	0
18	101.4	1.4	1.7	2	−3.4	0.0	0
19	97.2	−2.8	0.0	0	0.8	0.8	1
20	101.0	1.0	1.0	1	−3.0	0.0	0

When the tabular CUSUM indicates that the process is out of control, we should search for the assignable cause, take any corrective actions indicated, and restart the CUSUMs at zero. It may be helpful to have an estimate of the new process mean following the shift. This can be computed from

$$\hat{\mu} = \begin{cases} \mu_0 + K + \dfrac{s_H(i)}{n_H}, & \text{if } s_H(i) > H \\ \mu_0 - K - \dfrac{s_L(i)}{n_L}, & \text{if } s_L(i) > H \end{cases} \tag{15.33}$$

Also, an estimate of the time at which the assignable cause occurred is often taken as the sample time at which the upper or lower CUSUM (whichever one signaled) was last equal to zero.

It is also useful to present a graphical display of the tabular CUSUMs, which are sometimes called CUSUM status charts. They are constructed by plotting $s_H(i)$ and $s_L(i)$ versus the sample number. Figure 15.23 shows the CUSUM status chart for the data in Example 15.7. Each vertical bar represents the value of $s_H(i)$ and $s_L(i)$ in period i. With the decision interval plotted on the chart, the CUSUM status chart resembles a Shewhart control chart. We have also plotted the sample statistics x_i for each period on the CUSUM status chart as the solid dots. This frequently helps the user of the control chart to visualize the actual process performance that has led to a particular value of the CUSUM.

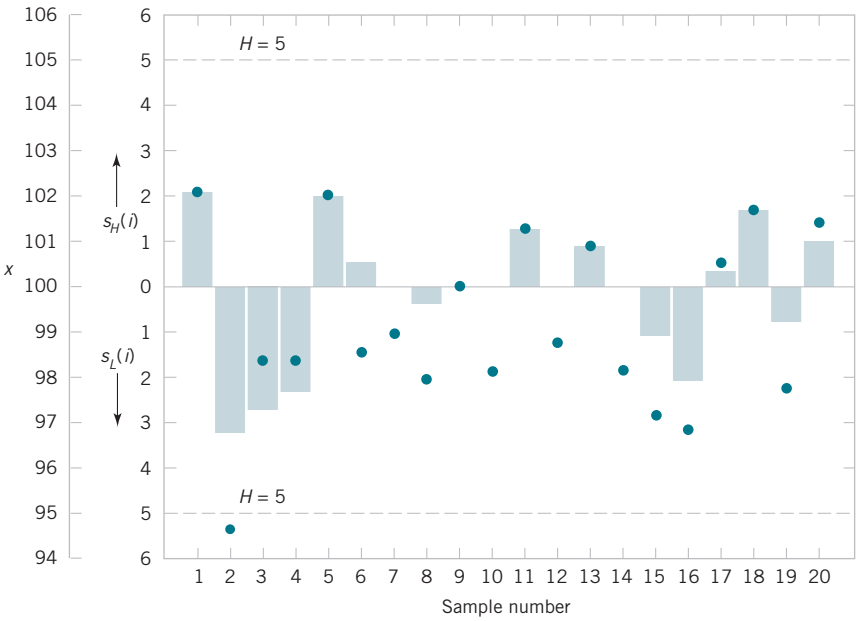


FIGURE 15.23

The CUSUM status chart for Example 15.7.

The tabular CUSUM is designed by choosing values for the reference value K and the decision interval H. We recommend that these parameters be selected to provide good average run-length values. There have been many analytical studies of CUSUM ARL performance. Based on them, we may give some general recommendations for selecting H and K. Define $H = h\sigma_{\overline{X}}$ and $K = k\sigma_{\overline{X}}$ where $\sigma_{\overline{X}}$ is the standard deviation of the sample variable used in forming the CUSUM (if $n = 1, \sigma_{\overline{X}} = \sigma_X$). Using $h = 4$ or $h = 5$ and $k = 1/2$ generally provide a CUSUM that has good ARL properties against a shift of about $1\sigma_{\overline{X}}$ (or $1\sigma_X$) in the process mean. If much larger or smaller shifts are of interest, set $k = \delta/2$ where δ is the size of the shift in standard deviation units.

To illustrate how well the recommendations of $h = 4$ or $h = 5$ with $k = 1/2$ work, consider these average run lengths in Table 15.11. Notice that a shift of $1\sigma_{\overline{x}}$ would be detected in either 8.38 samples (with $k = 1/2$ and $h = 4$) or 10.4 samples (with $k = 1/2$ and $h = 5$). By comparison,

TABLE 15.11 **Average Run Lengths for a CUSUM Control Chart with $k = 1/2$**

Shift in Mean (multiple of $\sigma_{\overline{X}}$)	$h = 4$	$h = 5$
0	168	465
0.25	74.2	139
0.50	26.6	38
0.75	13.3	17
1.00	8.38	10.4
1.50	4.75	5.75
2.00	3.34	4.01
2.50	2.62	3.11
3.00	2.19	2.57
4.00	1.71	2.01

Table 15.8 shows that an $\overline{X}$ chart would require approximately 43.9 samples, on the average, to detect this shift.

These design rules were used for the CUSUM in Example 15.7. We assumed that the process standard deviation $\sigma = 2$. (This is a reasonable value; see Example 15.2.) Then with $k = 1/2$ and $h = 5$, we would use

$$K = k\sigma = \frac{1}{2}(2) = 1 \quad \text{and} \quad H = h\sigma = 5(2) = 10$$

in the tabular CUSUM procedure.

Finally, we should note that supplemental procedures such as the Western Electric rules cannot be safely applied to the CUSUM because successive values of $S_H(i)$ and $S_L(i)$ are not independent. In fact, the CUSUM can be thought of as a weighted average, where the weights are stochastic or random. In effect, all CUSUM values are highly correlated, thereby causing the Western Electric rules to produce too many false alarms.

15.9 Other SPC Problem-Solving Tools

Although the control chart is a very powerful tool for investigating the causes of variation in a process, it is most effective when used with other SPC problem-solving tools. In this section, we illustrate some of these tools, using the printed circuit board defect data in Example 15.5.

Figure 15.17 shows a U chart for the number of defects in samples of five printed circuit boards. The chart exhibits statistical control, but the number of defects must be reduced. The average number of defects per board is $8/5 = 1.6$, and this level of defects would require extensive rework.

The first step in solving this problem is to construct a **Pareto diagram** of the individual defect types. The Pareto diagram shown in Figure 15.24, indicates that insufficient solder and solder balls are the most frequently occurring defects, accounting for $(109/160)100 = 68\%$ of the observed defects. Furthermore, the first five defect categories on the Pareto chart are all solder-related defects. This points to the flow solder process as a potential opportunity for improvement.

To improve the surface mount process, a team consisting of the operator, the shop supervisor, the manufacturing engineer responsible for the process, and a quality engineer meets to

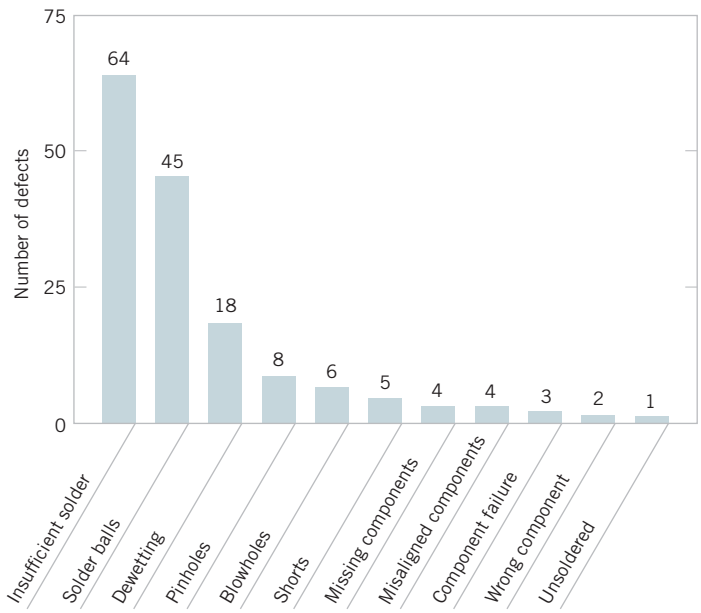


FIGURE 15.24
Pareto diagram for printed circuit board defects.

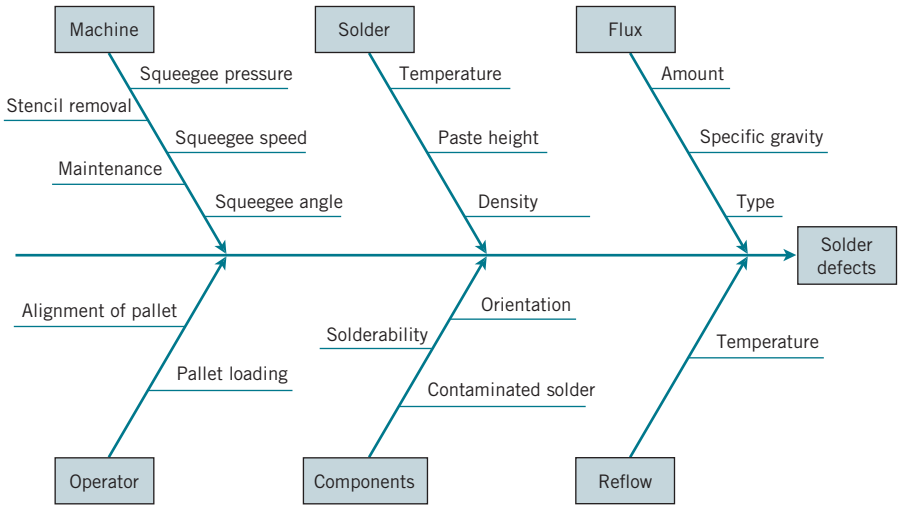


FIGURE 15.25

Cause-and-effect diagram for the printed circuit board flow solder process.

study potential causes of solder defects. They conduct a brainstorming session and produce the **cause-and-effect diagram** in Figure 15.25. The cause-and-effect diagram is widely used to display the various potential causes of defects in products and their interrelationships. They are useful in summarizing knowledge about the process.

As a result of the brainstorming session, the team tentatively identifies the following variables as potentially influential in creating solder defects: flux specific gravity, reflow temperature, squeegee speed, squeegee angle, paste height, reflow temperature, and board loading method. A statistically designed experiment could be used to investigate the effect of these seven variables on solder defects.

In addition, the team constructed a **defect concentration diagram** for the product; it is just a sketch or drawing of the product with the most frequently occurring defects shown on the part. This diagram is used to determine whether defects occur in the same location on the part. The defect concentration diagram for the printed circuit board is shown in Figure 15.26. This diagram indicates that most of the insufficient solder defects are near the front edge of the board. Further investigation showed that one of the pallets used to carry the boards was bent, causing the front edge of the board to make poor contact with the squeegee.

When the defective pallet was replaced, a designed experiment was used to investigate the seven variables discussed earlier. The results of this experiment indicated that several of these factors were influential and could be adjusted to reduce solder defects. After the results of the experiment were implemented, the percentage of solder joints requiring rework was reduced from 1% to under 100 parts per million (0.01%).

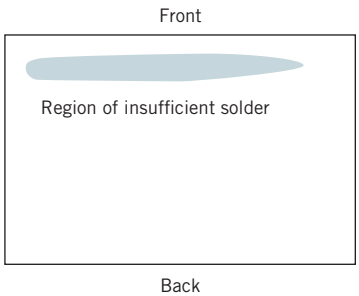


FIGURE 15.26

Defect concentration diagram for a printed circuit board.

15.10 Decision Theory

Quality improvement requires decision making in the presence of uncertainty as do many other management and engineering decisions. Consequently, a framework to characterize the decision problem in terms of actions, possible states, and associated probabilities is useful to quantitatively compare alternatives. **Decision theory** is the study of mathematical models for decision making. Actions are evaluated and selected based on the model and quantitative criteria.

15.10.1 Decision Models

A simple way to characterize decisions is in terms of actions, states with probabilities, and outcomes as costs (or profits). A decision usually involves a set of possible actions

$$A = \{a_1, a_2, \ldots, a_K\}$$

For example, one might purchase an extended warranty with the purchase of a vehicle (action a_1) or not (action a_2).

The possible future situations are represented with a collection of states

$$S = \{s_1, s_2, \ldots, s_M\}$$

One state occurs, but we are not certain of the future state at the time of our decision. For example, a possible state is that a major repair is required during the extended warranty; another is that no repair is required.

We often associate a probability, for example, p_m with each state s_m, so that $p_1 + p_2 + \ldots + p_M = 1$. The probabilities are important and can be difficult to estimate. Sometimes we have historical data from past performance and can derive estimates. In other cases, we might have to rely on the subjective belief of a collection of experts.

The outcome is often expressed in terms of economic cost (or profit) that depends on the actions and the state that occurs. That is, let

$$C_{km} = \text{cost when action } a_k \text{ is selected and state } s_m \text{ occurs}$$

For example, if we purchases an extended warranty and no repair is required, our loss is only the cost of the warranty. Clearly, if we do not purchase the warranty and a major repair is required, our loss is the cost of the repair.

EXAMPLE 15.8 | Extended Warranty Decision Problem

For deciding whether to purchase an extended warranty on a vehicle, we use the following model. The actions are

a_1 = purchase extended warranty

a_2 = do not purchase extended warranty

Assume that one of three states corresponding to a major, minor, or no repair can occur during the warranty period. We obtain probability estimates for each state. The states and associated probabilities are

s_1 = major repair, probability 0.1

s_2 = minor repair, probability 0.5

s_3 = no repair, probability 0.4

Finally, the costs C_{km} can be presented in a decision evaluation table in which each row is an action and each column is a state. We assume that the extended warranty coverage costs $200. Table 15.12 formally relates the cost of each action and possible future state.

TABLE 15.12 Decision Evaluation Table

Probabilities		0.1	0.5	0.4
States		s_1	s_2	s_3
Actions	a_1	$200	$200	$200
	a_2	$1,200	$300	$0

15.10.2 Decision Criteria

The numerical summary of the decision problem is presented with actions, states, and probabilities as illustrated in Example 15.8. However, the question regarding the "best" action still needs to be answered. Because the state is not known at the time of the decision, different costs are possible. Different criteria based on these costs (and the associated probabilities) can be used to select an action, but, as we show, these criteria do not always lead to the same action.

We might be pessimistic and focus on the worst possible state for each action. In this approach, we compares actions based on the maximum cost that can occur. For example, from Table 15.12, the maximum cost that occurs for action a_1 is $\max_m C_{1m} = \$200$ (regardless of the state). For action a_2, we have $\max_m C_{2m} = \max\{1200, 300, 0\} = \$1200$. A reasonable criterion is to select the action that minimizes this maximum cost and, in this case, the choice is a_1. In general, this approach selects the action a_k to minimize $\max_m C_{km}$.

Minimax Criterion

The **minimax** criterion selects the action a_k that corresponds to

$$\min_k \max_m C_{km} \tag{15.34}$$

The name clearly follows from the minimum and maximum that are computed.

The minimax criterion focuses on the worst-case (pessimistic) scenario and ignores the probabilities associated with the states. A state with a high cost for a specific action, even if the state is very unlikely, can penalize and eliminate the action.

An alternative criterion is to focus on the best-case (optimistic) scenario among the states and order the actions based on minimum cost $\min_m C_{km}$.

Minimin Criterion

The **minimin** criterion selects the action a_k that corresponds to

$$\min_k \min_m C_{km} \tag{15.35}$$

For the warranty example, $\min_m C_{1m} = 200$ and $\min_m C_{2m} = 0$ so that a_2 is selected with this criterion.

The previous criteria ignore the probabilities associated with the states. The most probable criterion evaluates an action based on the cost of the state with the most likely probability.

Most Probable Criterion

The **most probable** criterion selects the action to minimize the cost of the most probable state. (15.36)

For the warranty example, the most probable state is a minor repair. Based on this state alone, the cost associated with a_1 is \$200, and the cost associated with a_2 is \$300. Consequently, a_1 is chosen with this criterion.

An obvious criterion is the expected cost. Here we associate a random variable X_k with each action a_k. The distribution of the discrete random variable X_k consists of the costs and the associated probabilities for action a_k. The expected cost of a_k is defined simply as the expected value $E(X_k)$.

Expected Cost Criterion

The **expected cost** criterion selects that state to minimize the expected cost. (15.37)

For the warranty example, $E(X_1) = 200$ and $E(X_2) = 1200(0.1) + 300(0.5) + 0(0.4) = 270$. Consequently, the minimum expected cost is produced by action a_1.

A decision problem is often represented with a graph known as a **decision tree**. See Figure 15.27 for a decision tree of the warranty example. A square denotes a decision node where an action is selected. Each arc from a decision node represents an action. Each action arc terminates with a circular node to indicate that a state has been chosen (outside the control of the decision maker), and the states are represented by the arcs from the circular nodes. Each arc is labeled with the probability of the state. The cost is shown at the end of these arcs. This is the basic structure of the decision problem. Figure 15.27 computed from Equation 15.37 also shows the expected cost above each circle. Based on the expected costs, action a_1 is clearly preferred. We could replace the expected cost with other criterion discussed in this section to summarize the alternative actions.

In more complex problems, there is a series of actions and states (represented as rectangles and circles, respectively, in the decision tree). Probabilities are associated with each state, and a cost is associated with each path through the tree. Figure 15.28 provides an illustration. Given a criterion, we start at the end of a path and apply the criterion to determine the action. We continue the analysis from the end of a path until we reach the initial action node of the tree. This is illustrated in the following example.

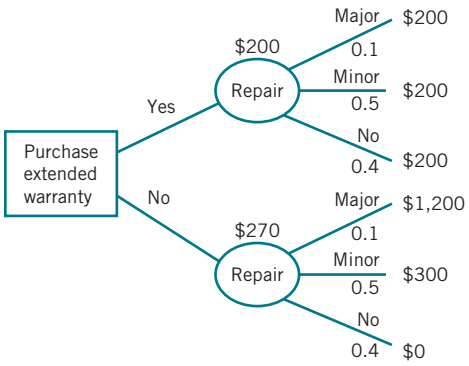


FIGURE 15.27

Decision tree for extended warranty example, expected cost of each action shown above circular modes.

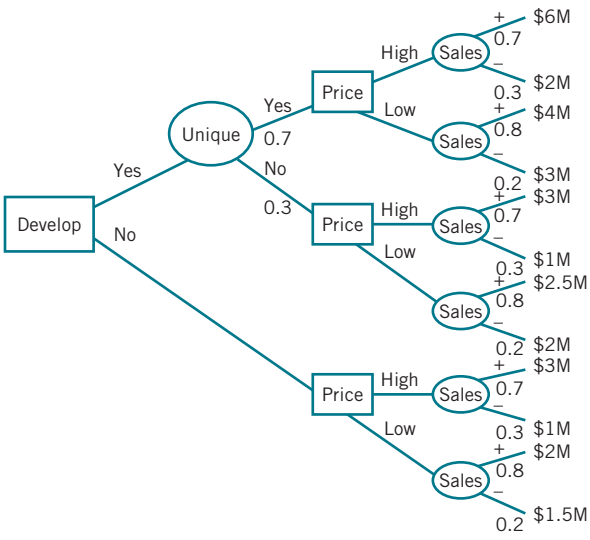


FIGURE 15.28

Decision tree for the develop or contract example.

EXAMPLE 15.9 | Develop or Contract Decision Problem

In this example, the basic decision task is extended to several decisions. The decision tree is shown in Figure 15.28. The first decision is whether to develop a new product or contract with a supplier. This is indicated by the box labeled *Develop?* If a new product is developed, it may be unique, but it may be more typical of what is currently available on the market. This is indicated by the circle labeled *Unique*? For either a new product or a contracted one, the price needs to be set. Here the decision is indicated by *Price?* boxes. The choices are either *high* or *low*. Finally, the market conditions when the product is available may be favorable or unfavorable to sales as indicated by the circle labeled *Sales*. Favorable and unfavorable markets are indicated by the arcs labeled + and −, respectively.

The probability of an arc is denoted by the number below it. For example, the probability that a unique product is developed is 0.7. Similarly, the probabilities of favorable or unfavorable markets are shown with the corresponding probabilities. Note that a lower price decision leads to the higher probability of a favorable market. Furthermore, the dollar amount shown in the figure indicates the profit to the corporation for the corresponding path through the decision tree. As mentioned previously, one might base a decision on profits (more generally, gains) rather than costs. In such a case, the objective is to maximize profits.

We can extend the procedure for a single decision node as follows. Start with the dollar amounts at the terminal nodes and work backward through the tree to evaluate a decision based on one of the criteria. Because we work with profits in this example, the pessimistic approach is to select the decision to maximize the minimum profit. For example, suppose that a new product is developed, the result is unique, and the price is set high. The two possible dollar values are \$6M and \$2M. The pessimistic approach is to value the decision to set the price high in this path as \$2M. Similarly, the decision to set the price low is valued at \$3M. Consequently, the decision along this path is to set the price low with a worst-case profit of \$3M.

Similarly, suppose that a new product is developed, the result is not unique, and the price is set high. The two possible dollar values are \$3M and \$1M. The pessimistic approach is to value the decision to set the price high in this path as \$1M. The decision to set the price low is pessimistically valued at \$2M. Consequently, the decision to develop a new product can result in a unique product (which is pessimistically evaluated as \$2M with probability 0.7) or a nonunique product (which is pessimistically evaluated at \$1M). The pessimistic view is that the decision to develop a new product generates a profit of \$1M.

Furthermore, suppose that the product is not developed (but contracted) and the price is set high with the pessimistic profit of \$1M. If the price is set low, the pessimistic profit is \$1.5M. Consequently, the price decision based on this criterion is to set the price low with a pessimistic profit of \$1.5M. Finally, the pessimistic profit from the decision to develop or not develop a new product is \$1M and \$1.5M, respectively. Therefore, based on this criterion, a new product is not developed.

Note that the probabilities do not enter into this decision. This was mentioned previously as one of the disadvantages of the pessimistic criterion. Alternative criteria are left as exercises.

15.11 Implementing SPC

The methods of statistical process control can provide significant payback to those companies that can successfully implement them. Although SPC seems to be a collection of statistically based problem-solving tools, there is more to its successful use than simply learning and using these tools. Management involvement and commitment to the quality-improvement process is the most vital component of SPC's potential success. Management is a role model, and others in the organization look to management for guidance and as an example. A team approach is also important, for it is usually difficult for one person alone to introduce process improvements. Many of the "magnificent seven" problem-solving tools are helpful in building an improvement team, including cause-and-effect diagrams, Pareto charts, and defect concentration diagrams. The basic SPC problem-solving tools must become widely known and widely used throughout the organization. Continuous training in SPC and quality improvement is necessary to achieve this widespread knowledge of the tools.

The objective of an SPC-based quality-improvement program is continuous improvement on a weekly, quarterly, and annual basis. SPC is not a one-time program to be applied when

the business is in trouble and later abandoned. Quality improvement must become part of the organizations culture.

The control chart is an important tool for process improvement. Processes do not naturally operate in an in-control state, and the use of control charts is an important step that must be taken early in an SPC program to eliminate assignable causes, reduce process variability, and stabilize process performance. To improve quality and productivity, we must begin to manage with facts and data, not just rely on judgment. Control charts are an important part of this change in management approach.

In implementing a companywide SPC program, we have found that the following elements are usually present in all successful efforts:

1. Management leadership
2. Team approach
3. Education of employees at all levels
4. Emphasis on continuous improvement
5. Mechanism for recognizing success

We cannot overemphasize the importance of management leadership and the team approach. Successful quality improvement is a "top-down" management-driven activity. It is also important to measure progress and success and to spread knowledge of this success throughout the organization. Communicating successful improvements throughout the company can provide motivation and incentive to improve other processes and to make continuous improvement a normal part of the way of doing business.

The philosophy of **W. Edwards Deming** provides an important framework for implementing quality and productivity improvement. Deming's philosophy is summarized in his 14 points for management. The adherence to these management principles has been an important factor in Japan's industrial success and continues to be the catalyst in that nation's quality- and productivity-improvement efforts. This philosophy has also now spread rapidly in the West. **Deming's 14 points** are as follows.

1. **Create a constancy of purpose focused on the improvement of products and services.** Constantly try to improve product design and performance. Investment in research, development, and innovation will have a long-term payback to the organization.
2. **Adopt a new philosophy of rejecting poor workmanship, defective products, or bad service.** It costs as much to produce a defective unit as it does to produce a good one (and sometimes more). The cost of dealing with scrap, rework, and other losses created by defectives is an enormous drain on company resources.
3. **Do not rely on mass inspection to "control quality."** All inspection can do is sort out defectives, and at this point, it is too late because we have already paid to produce these defectives. Inspection occurs too late in the process, is expensive, and is often ineffective. Quality results from the prevention of defectives through process improvement, not inspection.
4. **Do not award business to suppliers on the basis of price alone, but also consider quality.** Price is a meaningful measure of a supplier's product only if it is considered in relation to a measure of quality. In other words, the total cost of the item, not just the purchase price, must be considered. When quality is considered, the lowest bidder is frequently not the low-cost supplier. Preference should be given to suppliers who use modern methods of quality improvement in their business and who can demonstrate process control and capability.

5. **Focus on continuous improvement.** Constantly try to improve the production and service system. Involve the workforce in these activities and use statistical methods, particularly the SPC problem-solving tools discussed previously.
6. **Practice modern training methods and invest in training for all employees.** All employees should be trained in the technical aspects of their job, as well as in modern quality- and productivity-improvement methods. The training should encourage all employees to practice these methods every day.
7. **Practice modern supervision methods.** Supervision should not consist merely of passive surveillance of workers but also should be focused on helping the employees improve the system in which they work. The first goal of supervision should be to improve the work system and the product.
8. **Drive out fear.** Many workers are afraid to ask questions, report problems, or point out conditions that are barriers to quality and effective production. In many organizations, the economic loss associated with fear is large; only management can eliminate fear.
9. **Break down the barriers between functional areas of the business.** Teamwork among different organizational units is essential for effective quality and productivity improvement to take place.
10. **Eliminate targets, slogans, and numerical goals for the workforce.** A target such as "zero defects" is useless without a plan as to how to achieve it. In fact, these slogans and "programs" are usually counterproductive. Work to improve the system and provide information on that.
11. **Eliminate numerical quotas and work standards.** These standards have historically been set without regard to quality. Work standards are often symptoms of management's inability to understand the work process and to provide an effective management system focused on improving this process.
12. **Remove the barriers that discourage employees from doing their jobs.** Management must listen to employee suggestions, comments, and complaints. The person who is doing the job is the one who knows the most about it, and usually has valuable ideas about how to make the process work more effectively. The workforce is an important participant in the business, not just an opponent in collective bargaining.
13. **Institute an ongoing program of training and education for all employees.** Education in simple, powerful statistical techniques should be mandatory for all employees. Use of the basic SPC problem-solving tools, particularly the control chart, should become widespread in the business. As these charts become widespread, and as employees understand their uses, they are more likely to look for the causes of poor quality and to identify process improvements. Education is a way of making everyone partners in the quality-improvement process.
14. **Create a structure in top management that vigorously advocates the first 13 points.**

As we read Deming's 14 points, we notice two things. First, there is a strong emphasis on change. Second, the role of management in guiding this change process is of dominating importance. But what should be changed, and how should this change process be started? For example, if we want to improve the yield of a semiconductor manufacturing process, what should we do? It is in this area that statistical methods most frequently come into play. To improve the semiconductor process, we must determine which controllable factors in the process influence the number of defective units produced. To answer this question, we must collect data on the process and see how the system reacts to changes in the process variables. Statistical methods, including the SPC and experimental design techniques in this book, can contribute to this knowledge.

Important Terms and Concepts

Assignable causes
Attributes control charts
Average run length (ARL)
C chart
Cause-and-effect diagram
Center line
Chance causes
Control chart
Control limits
Cumulative sum control chart (CUSUM)
Decision theory
Decision tree
Defect concentration diagram
Deming's 14 points
Expected cost criterion
Exponentially weighted moving-average control chart (EWMA)
False alarm
Fraction-defective control chart
Individuals control chart (*X* chart)
Minimax criterion
Minimin criterion
Most probable criterion
Moving range
Natural tolerance limits
NP chart
P chart
Pareto diagram
Problem-solving tools
Process capability
Process capability ratio (PCR, PCR_k)
Quality improvement
R chart
Rational subgroup
Run rules
S chart
Shewhart control chart
6-sigma process
Specification limits
Statistical process control (SPC)
Statistical quality control
U chart
Variables control charts
Warning limits
Western Electric rules
$\overline{X}$ chart

Appendices

Appendix A	**Statistical Tables and Charts**	**A-3**
Table I	Summary of Common Probability Distributions	A-4
Table II	Cumulative Binomial Probabilities $P(X \leq x)$	A-5
Table III	Cumulative Standard Normal Distribution	A-8
Table IV	Percentage Points $\chi^2_{\alpha,\nu}$ of the Chi-Squared Distribution	A-10
Table V	Percentage Points $t_{\alpha,\nu}$ of the t Distribution	A-11
Table VI	Percentage Points f_{α,ν_1,ν_2} of the F Distribution	A-12
Chart VII	Operating Characteristic Curves	A-17
Table VIII	Critical Values for the Sign Test	A-26
Table IX	Critical Values for the Wilcoxon Signed-Rank Test	A-26
Table X	Critical Values for the Wilcoxon Rank-Sum Test	A-27
Table XI	Factors for Constructing Variables Control Charts	A-28
Table XII	Factors for Tolerance Intervals	A-29
Appendix B	**Bibliography**	**A-31**

Appendix **A**

Statistical Tables and Charts

TABLE I **Summary of Common Probability Distributions**

Name	Probability Distribution	Mean	Variance	Section in Book
Discrete				
Uniform	$\frac{1}{n}, a \le b$	$\frac{(b+a)}{2}$	$\frac{(b-a+1)^2-1}{12}$	3-5
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$	np	$np(1-p)$	3-6
	$x = 0, 1, \ldots, n, 0 \le p \le 1$			
Geometric	$(1-p)^{x-1}p$ $x = 1, 2, \ldots, 0 \le p \le 1$	$1/p$	$(1-p)/p^2$	3-7
Negative binomial	$\binom{x-1}{r-1}(1-p)^{x-r}p^r$ $x = r, r+1, r+2, \ldots, 0 \le p \le 1$	r/p	$r(1-p)/p^2$	3-7
Hypergeometric	$\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$ $x = \max(0, n-N+K), 1, \ldots$ $\min(K, n), K \le N, n \le N$	np where $p = \frac{K}{N}$	$np(1-p)\left(\frac{N-n}{N-1}\right)$	3-8
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots, 0 < \lambda$	λ	λ	3-9
Continuous				
Uniform	$\frac{1}{b-a}, a \le x \le b$	$\frac{(b+a)}{2}$	$\frac{(b-a)^2}{12}$	4-5
Normal	$\frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}$ $-\infty < x < \infty, -\infty < \mu < \infty, 0 < \sigma$	μ	σ^2	4-6
Exponential	$\lambda e^{-\lambda x}, 0 \le x, 0 < \lambda$	$1/\lambda$	$1/\lambda^2$	4-8
Erlang	$\frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!}, 0 < x, r = 1, 2, \ldots$	r/λ	r/λ^2	4-9.1
Gamma	$\frac{\lambda x^{r-1} e^{-\lambda x}}{\Gamma(r)}, 0 < x, 0 < r, 0 < \lambda$	r/λ	r/λ^2	4-9.2
Weibull	$\frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} e^{-(x/\delta)^\beta}$ $0 < x, 0 < \beta, 0 < \delta$	$\delta\Gamma\left(1+\frac{1}{\beta}\right)$	$\delta^2\Gamma\left(1+\frac{2}{\beta}\right) - \delta^2\left[\Gamma\left(1+\frac{1}{\beta}\right)\right]^2$	4-10
Lognormal	$\frac{1}{x\omega\sqrt{2\pi}} \exp\left(\frac{-[\ln(x)-\theta]^2}{2\omega^2}\right)$	$e^{\theta+\omega^2/2}$	$e^{2\theta+\omega^2}(e^{\omega^2}-1)$	4-11
Beta	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$ $0 \le x \le 1, 0 < \alpha, 0 < \beta$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	4-12

TABLE II **Cumulative Binomial Probabilities $P(X \leq x)$**

		P										
n	x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
1	0	0.9000	0.8000	0.7000	0.6000	0.5000	0.4000	0.3000	0.2000	0.1000	0.0500	0.0100
2	0	0.8100	0.6400	0.4900	0.3600	0.2500	0.1600	0.0900	0.0400	0.0100	0.0025	0.0001
	1	0.9900	0.9600	0.9100	0.8400	0.7500	0.6400	0.5100	0.3600	0.1900	0.0975	0.0199
3	0	0.7290	0.5120	0.3430	0.2160	0.1250	0.0640	0.0270	0.0080	0.0010	0.0001	0.0000
	1	0.9720	0.8960	0.7840	0.6480	0.5000	0.3520	0.2160	0.1040	0.0280	0.0073	0.0003
	2	0.9990	0.9920	0.9730	0.9360	0.8750	0.7840	0.6570	0.4880	0.2710	0.1426	0.0297
4	0	0.6561	0.4096	0.2401	0.1296	0.0625	0.0256	0.0081	0.0016	0.0001	0.0000	0.0000
	1	0.9477	0.8192	0.6517	0.4752	0.3125	0.1792	0.0837	0.0272	0.0037	0.0005	0.0000
	2	0.9963	0.9728	0.9163	0.8208	0.6875	0.5248	0.3483	0.1808	0.0523	0.0140	0.0006
	3	0.9999	0.9984	0.9919	0.9744	0.9375	0.8704	0.7599	0.5904	0.3439	0.1855	0.0394
5	0	0.5905	0.3277	0.1681	0.0778	0.0313	0.0102	0.0024	0.0003	0.0000	0.0000	0.0000
	1	0.9185	0.7373	0.5282	0.3370	0.1875	0.0870	0.0308	0.0067	0.0005	0.0000	0.0000
	2	0.9914	0.9421	0.8369	0.6826	0.5000	0.3174	0.1631	0.0579	0.0086	0.0012	0.0000
	3	0.9995	0.9933	0.9692	0.9130	0.8125	0.6630	0.4718	0.2627	0.0815	0.0226	0.0010
	4	1.0000	0.9997	0.9976	0.9898	0.6988	0.9222	0.8319	0.6723	0.4095	0.2262	0.0490
6	0	0.5314	0.2621	0.1176	0.0467	0.0156	0.0041	0.0007	0.0001	0.0000	0.0000	0.0000
	1	0.8857	0.6554	0.4202	0.2333	0.1094	0.0410	0.0109	0.0016	0.0001	0.0000	0.0000
	2	0.9842	0.9011	0.7443	0.5443	0.3438	0.1792	0.0705	0.0170	0.0013	0.0001	0.0000
	3	0.9987	0.9830	0.9295	0.8208	0.6563	0.4557	0.2557	0.0989	0.0159	0.0022	0.0000
	4	0.9999	0.9984	0.9891	0.9590	0.9806	0.7667	0.5798	0.3446	0.1143	0.0328	0.0015
	5	1.0000	0.9999	0.9993	0.9959	0.9844	0.9533	0.8824	0.7379	0.4686	0.2649	0.0585
7	0	0.4783	0.2097	0.0824	0.0280	0.0078	0.0016	0.0002	0.0000	0.0000	0.0000	0.0000
	1	0.8503	0.5767	0.3294	0.1586	0.0625	0.0188	0.0038	0.0004	0.0000	0.0000	0.0000
	2	0.9743	0.8520	0.6471	0.4199	0.2266	0.0963	0.0288	0.0047	0.0002	0.0000	0.0000
	3	0.9973	0.9667	0.8740	0.7102	0.5000	0.2898	0.1260	0.0333	0.0027	0.0002	0.0000
	4	0.9998	0.9953	0.9712	0.9037	0.7734	0.5801	0.3529	0.1480	0.0257	0.0038	0.0000
	5	1.0000	0.9996	0.9962	0.9812	0.9375	0.8414	0.6706	0.4233	0.1497	0.0444	0.0020
	6	1.0000	1.0000	0.9998	0.9984	0.9922	0.9720	0.9176	0.7903	0.5217	0.3017	0.0679
8	0	0.4305	0.1678	0.0576	0.0168	0.0039	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000
	1	0.8131	0.5033	0.2553	0.1064	0.0352	0.0085	0.0013	0.0001	0.0000	0.0000	0.0000
	2	0.9619	0.7969	0.5518	0.3154	0.1445	0.0498	0.0113	0.0012	0.0000	0.0000	0.0000
	3	0.9950	0.9437	0.8059	0.5941	0.3633	0.1737	0.0580	0.0104	0.0004	0.0000	0.0000
	4	0.9996	0.9896	0.9420	0.8263	0.6367	0.4059	0.1941	0.0563	0.0050	0.0004	0.0000
	5	1.0000	0.9988	0.9887	0.9502	0.8555	0.6846	0.4482	0.2031	0.0381	0.0058	0.0001
	6	1.0000	0.9999	0.9987	0.9915	0.9648	0.8936	0.7447	0.4967	0.1869	0.0572	0.0027
	7	1.0000	1.0000	0.9999	0.9993	0.9961	0.9832	0.9424	0.8322	0.5695	0.3366	0.0773
9	0	0.3874	0.1342	0.0404	0.0101	0.0020	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.7748	0.4362	0.1960	0.0705	0.0195	0.0038	0.0004	0.0000	0.0000	0.0000	0.0000
	2	0.9470	0.7382	0.4628	0.2318	0.0889	0.0250	0.0043	0.0003	0.0000	0.0000	0.0000
	3	0.9917	0.9144	0.7297	0.4826	0.2539	0.0994	0.0253	0.0031	0.0001	0.0000	0.0000
	4	0.9991	0.9804	0.9012	0.7334	0.5000	0.2666	0.0988	0.0196	0.0009	0.0000	0.0000
	5	0.9999	0.9969	0.9747	0.9006	0.7461	0.5174	0.2703	0.0856	0.0083	0.0006	0.0000
	6	1.0000	0.9997	0.9957	0.9750	0.9102	0.7682	0.5372	0.2618	0.0530	0.0084	0.0001
	7	1.0000	1.0000	0.9996	0.9962	0.9805	0.9295	0.8040	0.5638	0.2252	0.0712	0.0034
	8	1.0000	1.0000	1.0000	0.9997	0.9980	0.9899	0.9596	0.8658	0.6126	0.3698	0.0865

TABLE II **Cumulative Binomial Probabilities $P(X \leq x)$ (*continued*)**

		P										
n	***x***	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
10	0	0.3487	0.1074	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.7361	0.3758	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000	0.0000	0.0000
	2	0.9298	0.6778	0.3828	0.1673	0.0547	0.0123	0.0016	0.0001	0.0000	0.0000	0.0000
	3	0.9872	0.8791	0.6496	0.3823	0.1719	0.0548	0.0106	0.0009	0.0000	0.0000	0.0000
	4	0.9984	0.9672	0.8497	0.6331	0.3770	0.1662	0.0473	0.0064	0.0001	0.0000	0.0000
	5	0.9999	0.9936	0.9527	0.8338	0.6230	0.3669	0.1503	0.0328	0.0016	0.0001	0.0000
	6	1.0000	0.9991	0.9894	0.9452	0.8281	0.6177	0.3504	0.1209	0.0128	0.0010	0.0000
	7	1.0000	0.9999	0.9984	0.9877	0.9453	0.8327	0.6172	0.3222	0.0702	0.0115	0.0001
	8	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.6242	0.2639	0.0861	0.0043
	9	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.8926	0.6513	0.4013	0.0956
11	0	0.3138	0.0859	0.0198	0.0036	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.6974	0.3221	0.1130	0.0302	0.0059	0.0007	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.9104	0.6174	0.3127	0.1189	0.0327	0.0059	0.0006	0.0000	0.0000	0.0000	0.0000
	3	0.9815	0.8389	0.5696	0.2963	0.1133	0.0293	0.0043	0.0002	0.0000	0.0000	0.0000
	4	0.9972	0.9496	0.7897	0.5328	0.2744	0.0994	0.0216	0.0020	0.0000	0.0000	0.0000
	5	0.9997	0.9883	0.9218	0.7535	0.5000	0.2465	0.0782	0.0117	0.0003	0.0000	0.0000
	6	1.0000	0.9980	0.9784	0.9006	0.7256	0.4672	0.2103	0.0504	0.0028	0.0001	0.0000
	7	1.0000	0.9998	0.9957	0.9707	0.8867	0.7037	0.4304	0.1611	0.0185	0.0016	0.0000
	8	1.0000	1.0000	0.9994	0.9941	0.9673	0.8811	0.6873	0.3826	0.0896	0.0152	0.0002
	9	1.0000	1.0000	1.0000	0.9993	0.9941	0.9698	0.8870	0.6779	0.3026	0.1019	0.0052
	10	1.0000	1.0000	1.0000	1.0000	0.9995	0.9964	0.9802	0.9141	0.6862	0.4312	0.1047
12	0	0.2824	0.0687	0.0138	0.0022	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.6590	0.2749	0.0850	0.0196	0.0032	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.8891	0.5583	0.2528	0.0834	0.0193	0.0028	0.0002	0.0000	0.0000	0.0000	0.0000
	3	0.9744	0.7946	0.4925	0.2253	0.0730	0.0153	0.0017	0.0001	0.0000	0.0000	0.0000
	4	0.9957	0.9274	0.7237	0.4382	0.1938	0.0573	0.0095	0.0006	0.0000	0.0000	0.0000
	5	0.9995	0.9806	0.8822	0.6652	0.3872	0.1582	0.0386	0.0039	0.0001	0.0000	0.0000
	6	0.9999	0.9961	0.9614	0.8418	0.6128	0.3348	0.1178	0.0194	0.0005	0.0000	0.0000
	7	1.0000	0.9994	0.9905	0.9427	0.8062	0.5618	0.2763	0.0726	0.0043	0.0002	0.0000
	8	1.0000	0.9999	0.9983	0.9847	0.9270	0.7747	0.5075	0.2054	0.0256	0.0022	0.0000
	9	1.0000	1.0000	0.9998	0.9972	0.9807	0.9166	0.7472	0.4417	0.1109	0.0196	0.0002
	10	1.0000	1.0000	1.0000	0.9997	0.9968	0.9804	0.9150	0.7251	0.3410	0.1184	0.0062
	11	1.0000	1.0000	1.0000	1.0000	0.9998	0.9978	0.9862	0.9313	0.7176	0.4596	0.1136
13	0	0.2542	0.0550	0.0097	0.0013	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.6213	0.2336	0.0637	0.0126	0.0017	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.8661	0.5017	0.2025	0.0579	0.0112	0.0013	0.0001	0.0000	0.0000	0.0000	0.0000
	3	0.9658	0.7473	0.4206	0.1686	0.0461	0.0078	0.0007	0.0000	0.0000	0.0000	0.0000
	4	0.9935	0.9009	0.6543	0.3530	0.1334	0.0321	0.0040	0.0002	0.0000	0.0000	0.0000
	5	0.9991	0.9700	0.8346	0.5744	0.2905	0.0977	0.0182	0.0012	0.0000	0.0000	0.0000
	6	0.9999	0.9930	0.9376	0.7712	0.5000	0.2288	0.0624	0.0070	0.0001	0.0000	0.0000
	7	1.0000	0.9988	0.9818	0.9023	0.7095	0.4256	0.1654	0.0300	0.0009	0.0000	0.0000
	8	1.0000	0.9988	0.9960	0.9679	0.8666	0.6470	0.3457	0.0991	0.0065	0.0003	0.0000
	9	1.0000	1.0000	0.9993	0.9922	0.9539	0.8314	0.5794	0.2527	0.0342	0.0031	0.0000
	10	1.0000	1.0000	0.9999	0.9987	0.9888	0.9421	0.7975	0.4983	0.1339	0.0245	0.0003
	11	1.0000	1.0000	1.0000	0.9999	0.9983	0.9874	0.9363	0.7664	0.3787	0.1354	0.0072
	12	1.0000	1.0000	1.0000	1.0000	0.9999	0.9987	0.9903	0.9450	0.7458	0.4867	0.1225

TABLE II **Cumulative Binomial Probabilities $P(X \leq x)$ (*continued*)**

		P										
n	***x***	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
14	0	0.2288	0.0440	0.0068	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.5846	0.1979	0.0475	0.0081	0.0009	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.8416	0.4481	0.1608	0.0398	0.0065	0.0006	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.9559	0.6982	0.3552	0.1243	0.0287	0.0039	0.0002	0.0000	0.0000	0.0000	0.0000
	4	0.9908	0.8702	0.5842	0.2793	0.0898	0.0175	0.0017	0.0000	0.0000	0.0000	0.0000
	5	0.9985	0.9561	0.7805	0.4859	0.2120	0.0583	0.0083	0.0004	0.0000	0.0000	0.0000
	6	0.9998	0.9884	0.9067	0.6925	0.3953	0.1501	0.0315	0.0024	0.0000	0.0000	0.0000
	7	1.0000	0.9976	0.9685	0.8499	0.6047	0.3075	0.0933	0.0116	0.0002	0.0000	0.0000
	8	1.0000	0.9996	0.9917	0.9417	0.7880	0.5141	0.2195	0.0439	0.0015	0.0000	0.0000
	9	1.0000	1.0000	0.9983	0.9825	0.9102	0.7207	0.4158	0.1298	0.0092	0.0004	0.0000
	10	1.0000	1.0000	0.9998	0.9961	0.9713	0.8757	0.6448	0.3018	0.0441	0.0042	0.0000
	11	1.0000	1.0000	1.0000	0.9994	0.9935	0.9602	0.8392	0.5519	0.1584	0.0301	0.0003
	12	1.0000	1.0000	1.0000	0.9999	0.9991	0.9919	0.9525	0.8021	0.4154	0.1530	0.0084
	13	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9932	0.9560	0.7712	0.5123	0.1313
15	0	0.2059	0.0352	0.0047	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.5490	0.1671	0.0353	0.0052	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.8159	0.3980	0.1268	0.0271	0.0037	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.9444	0.6482	0.2969	0.0905	0.0176	0.0019	0.0001	0.0000	0.0000	0.0000	0.0000
	4	0.9873	0.8358	0.5155	0.2173	0.0592	0.0093	0.0007	0.0000	0.0000	0.0000	0.0000
	5	0.9978	0.9389	0.7216	0.4032	0.1509	0.0338	0.0037	0.0001	0.0000	0.0000	0.0000
	6	0.9997	0.9819	0.8689	0.6098	0.3036	0.0950	0.0152	0.0008	0.0000	0.0000	0.0000
	7	1.0000	0.9958	0.9500	0.7869	0.5000	0.2131	0.0500	0.0042	0.0000	0.0000	0.0000
	8	1.0000	0.9992	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.0003	0.0000	0.0000
	9	1.0000	0.9999	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.0022	0.0001	0.0000
	10	1.0000	1.0000	0.9993	0.9907	0.9408	0.7827	0.4845	0.1642	0.0127	0.0006	0.0000
	11	1.0000	1.0000	0.9999	0.9981	0.9824	0.9095	0.7031	0.3518	0.0556	0.0055	0.0000
	12	1.0000	1.0000	1.0000	0.9997	0.9963	0.9729	0.8732	0.6020	01841	0.0362	0.0004
	13	1.0000	1.0000	1.0000	1.0000	0.9995	0.9948	0.9647	0.8329	0.4510	0.1710	0.0096
	14	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9953	0.9648	0.7941	0.5367	0.1399
20	0	0.1216	0.0115	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3917	0.0692	0.0076	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.6769	0.2061	0.0355	0.0036	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.8670	0.4114	0.1071	0.0160	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.9568	0.6296	0.2375	0.0510	0.0059	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.9887	0.8042	0.4164	0.1256	0.0207	0.0016	0.0000	0.0000	0.0000	0.0000	0.0000
	6	0.9976	0.9133	0.6080	0.2500	0.0577	0.0065	0.0003	0.0000	0.0000	0.0000	0.0000
	7	0.9996	0.9679	0.7723	0.4159	0.1316	0.0210	0.0013	0.0000	0.0000	0.0000	0.0000
	8	0.9999	0.9900	0.8867	0.5956	0.2517	0.0565	0.0051	0.0001	0.0000	0.0000	0.0000
	9	1.0000	0.9974	0.9520	0.7553	0.4119	0.1275	0.0171	0.0006	0.0000	0.0000	0.0000
	10	1.0000	0.9994	0.9829	0.8725	0.5881	0.2447	0.0480	0.0026	0.0000	0.0000	0.0000
	11	1.0000	0.9999	0.9949	0.9435	0.7483	0.4044	0.1133	0.0100	0.0001	0.0000	0.0000
	12	1.0000	1.0000	0.9987	0.9790	0.8684	0.5841	0.2277	0.0321	0.0004	0.0000	0.0000
	13	1.0000	1.0000	0.9997	0.9935	0.9423	0.7500	0.3920	0.0867	0.0024	0.0000	0.0000
	14	1.0000	1.0000	1.0000	0.9984	0.9793	0.8744	0.5836	0.1958	0.0113	0.0003	0.0000
	15	1.0000	1.0000	1.0000	0.9997	0.9941	0.9490	0.7625	0.3704	0.0432	0.0026	0.0000
	16	1.0000	1.0000	1.0000	1.0000	0.9987	0.9840	0.8929	0.5886	0.1330	0.0159	0.0000
	17	1.0000	1.0000	1.0000	1.0000	0.9998	0.9964	0.9645	0.7939	0.3231	0.0755	0.0010
	18	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9924	0.9308	0.6083	0.2642	0.0169
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9992	0.9885	0.8784	0.6415	0.1821

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$

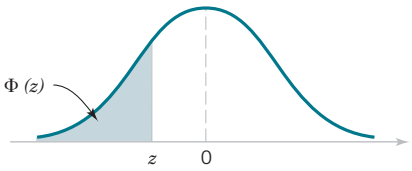


TABLE III **Cumulative Standard Normal Distribution**

z	−0.09	−0.08	−0.07	−0.06	−0.05	−0.04	−0.03	−0.02	−0.01	−0.00
−3.9	0.000033	0.000034	0.000036	0.000037	0.000039	0.000041	0.000042	0.000044	0.000046	0.000048
−3.8	0.000050	0.000052	0.000054	0.000057	0.000059	0.000062	0.000064	0.000067	0.000069	0.000072
−3.7	0.000075	0.000078	0.000082	0.000085	0.000088	0.000092	0.000096	0.000100	0.000104	0.000108
−3.6	0.000112	0.000117	0.000121	0.000126	0.000131	0.000136	0.000142	0.000147	0.000153	0.000159
−3.5	0.000165	0.000172	0.000179	0.000185	0.000193	0.000200	0.000208	0.000216	0.000224	0.000233
−3.4	0.000242	0.000251	0.000260	0.000270	0.000280	0.000291	0.000302	0.000313	0.000325	0.000337
−3.3	0.000350	0.000362	0.000376	0.000390	0.000404	0.000419	0.000434	0.000450	0.000467	0.000483
−3.2	0.000501	0.000519	0.000538	0.000557	0.000577	0.000598	0.000619	0.000641	0.000664	0.000687
−3.1	0.000711	0.000736	0.000762	0.000789	0.000816	0.000845	0.000874	0.000904	0.000935	0.000968
−3.0	0.001001	0.001035	0.001070	0.001107	0.001144	0.001183	0.001223	0.001264	0.001306	0.001350
−2.9	0.001395	0.001441	0.001489	0.001538	0.001589	0.001641	0.001695	0.001750	0.001807	0.001866
−2.8	0.001926	0.001988	0.002052	0.002118	0.002186	0.002256	0.002327	0.002401	0.002477	0.002555
−2.7	0.002635	0.002718	0.002803	0.002890	0.002980	0.003072	0.003167	0.003264	0.003364	0.003467
−2.6	0.003573	0.003681	0.003793	0.003907	0.004025	0.004145	0.004269	0.004396	0.004527	0.004661
−2.5	0.004799	0.004940	0.005085	0.005234	0.005386	0.005543	0.005703	0.005868	0.006037	0.006210
−2.4	0.006387	0.006569	0.006756	0.006947	0.007143	0.007344	0.007549	0.007760	0.007976	0.008198
−2.3	0.008424	0.008656	0.008894	0.009137	0.009387	0.009642	0.009903	0.010170	0.010444	0.010724
−2.2	0.011011	0.011304	0.011604	0.011911	0.012224	0.012545	0.012874	0.013209	0.013553	0.013903
−2.1	0.014262	0.014629	0.015003	0.015386	0.015778	0.016177	0.016586	0.017003	0.017429	0.017864
−2.0	0.018309	0.018763	0.019226	0.019699	0.020182	0.020675	0.021178	0.021692	0.022216	0.022750
−1.9	0.023295	0.023852	0.024419	0.024998	0.025588	0.026190	0.026803	0.027429	0.028067	0.028717
−1.8	0.029379	0.030054	0.030742	0.031443	0.032157	0.032884	0.033625	0.034379	0.035148	0.035930
−1.7	0.036727	0.037538	0.038364	0.039204	0.040059	0.040929	0.041815	0.042716	0.043633	0.044565
−1.6	0.045514	0.046479	0.047460	0.048457	0.049471	0.050503	0.051551	0.052616	0.053699	0.054799
−1.5	0.055917	0.057053	0.058208	0.059380	0.060571	0.061780	0.063008	0.064256	0.065522	0.066807
−1.4	0.068112	0.069437	0.070781	0.072145	0.073529	0.074934	0.076359	0.077804	0.079270	0.080757
−1.3	0.082264	0.083793	0.085343	0.086915	0.088508	0.090123	0.091759	0.093418	0.095098	0.096801
−1.2	0.098525	0.100273	0.102042	0.103835	0.105650	0.107488	0.109349	0.111233	0.113140	0.115070
−1.1	0.117023	0.119000	0.121001	0.123024	0.125072	0.127143	0.129238	0.131357	0.133500	0.135666
−1.0	0.137857	0.140071	0.142310	0.144572	0.146859	0.149170	0.151505	0.153864	0.156248	0.158655
−0.9	0.161087	0.163543	0.166023	0.168528	0.171056	0.173609	0.176185	0.178786	0.181411	0.184060
−0.8	0.186733	0.189430	0.192150	0.194894	0.197662	0.200454	0.203269	0.206108	0.208970	0.211855
−0.7	0.214764	0.217695	0.220650	0.223627	0.226627	0.229650	0.232695	0.235762	0.238852	0.241964
−0.6	0.245097	0.248252	0.251429	0.254627	0.257846	0.261086	0.264347	0.267629	0.270931	0.274253
−0.5	0.277595	0.280957	0.284339	0.287740	0.291160	0.294599	0.298056	0.301532	0.305026	0.308538
−0.4	0.312067	0.315614	0.319178	0.322758	0.326355	0.329969	0.333598	0.337243	0.340903	0.344578
−0.3	0.348268	0.351973	0.355691	0.359424	0.363169	0.366928	0.370700	0.374484	0.378281	0.382089
−0.2	0.385908	0.389739	0.393580	0.397432	0.401294	0.405165	0.409046	0.412936	0.416834	0.420740
−0.1	0.424655	0.428576	0.432505	0.436441	0.440382	0.444330	0.448283	0.452242	0.456205	0.460172
0.0	0.464144	0.468119	0.472097	0.476078	0.480061	0.484047	0.488033	0.492022	0.496011	0.500000

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$

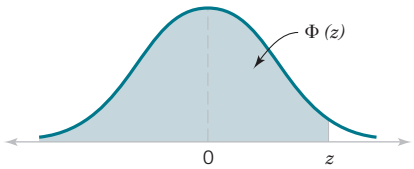


TABLE III **Cumulative Standard Normal Distribution (*continued*)**

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965621	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999533	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999650
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999821	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.999950
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967

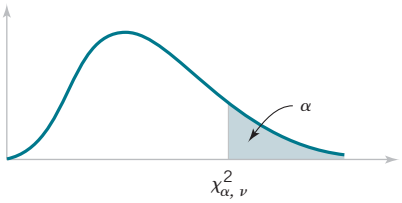


TABLE IV **Percentage Points $\chi^2_{\alpha,\nu}$ of the Chi-Squared Distribution**

ν \ α	.995	.990	.975	.950	.900	.500	.100	.050	.025	.010	.005
1	.00+	.00+	.00+	.00+	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	89.33	107.57	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

ν = degrees of freedom.

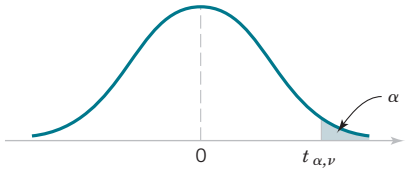


TABLE V **Percentage Points $t_{\alpha,\nu}$ of the t Distribution**

ν \ α	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.258	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	.256	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	.256	.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	.255	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	.254	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	.254	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	.253	.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

ν = degrees of freedom.

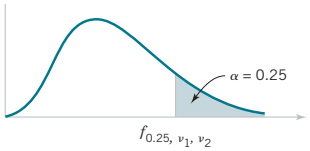


TABLE VI Percentage Points f_{α,v_1,v_2} of the F Distribution

$f_{0.25,v_1,v_2}$

v_2 \ v_1	Degrees of freedom for the numerator (v_1)																		
Degrees of freedom for the denominator (v_2)	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26	9.32	9.41	9.49	9.58	9.63	9.67	9.71	9.76	9.80	9.85
2	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38	3.39	3.41	3.43	3.43	3.44	3.45	3.46	3.47	3.48
3	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44	2.45	2.46	2.46	2.46	2.47	2.47	2.47	2.47	2.47
4	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
5	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.88	1.88	1.88	1.88	1.87	1.87	1.87
6	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77	1.77	1.76	1.76	1.75	1.75	1.75	1.74	1.74	1.74
7	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.70	1.69	1.68	1.68	1.67	1.67	1.66	1.66	1.65	1.65	1.65
8	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63	1.63	1.62	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.58
9	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.54	1.53	1.53
10	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.53	1.52	1.52	1.51	1.51	1.50	1.49	1.48
11	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.47	1.46	1.45
12	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.45	1.44	1.43	1.42
13	1.45	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.42	1.41	1.40
14	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.41	1.40	1.39	1.38
15	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.46	1.45	1.44	1.43	1.41	1.41	1.40	1.39	1.38	1.37	1.36
16	1.42	1.51	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.44	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34
17	1.42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33
18	1.41	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.32
19	1.41	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.33	1.32	1.30
20	1.40	1.49	1.48	1.47	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31	1.29
21	1.40	1.48	1.48	1.46	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28
22	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.39	1.37	1.36	1.34	1.33	1.32	1.31	1.30	1.29	1.28
23	1.39	1.47	1.47	1.45	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28	1.27
24	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.36	1.35	1.33	1.32	1.31	1.30	1.29	1.28	1.26
25	1.39	1.47	1.46	1.44	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.32	1.31	1.29	1.28	1.27	1.25
26	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.37	1.35	1.34	1.32	1.31	1.30	1.29	1.28	1.26	1.25
27	1.38	1.46	1.45	1.43	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.33	1.32	1.31	1.30	1.28	1.27	1.26	1.24
28	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.31	1.30	1.29	1.28	1.27	1.25	1.24
29	1.38	1.45	1.45	1.43	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.32	1.31	1.30	1.29	1.27	1.26	1.25	1.23
30	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.34	1.32	1.30	1.29	1.28	1.27	1.26	1.24	1.23
40	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.31	1.30	1.28	1.26	1.25	1.24	1.22	1.21	1.19
60	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.27	1.25	1.24	1.22	1.21	1.19	1.17	1.15
120	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.26	1.24	1.22	1.21	1.19	1.18	1.16	1.13	1.10
∞	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.22	1.19	1.18	1.16	1.14	1.12	1.08	1.00

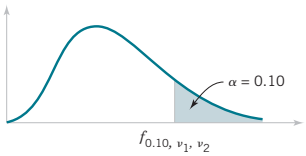


TABLE VI **Percentage Points f_{α,ν_1,ν_2} of the F Distribution (*continued*)**

$f_{0.10,\nu_1,\nu_2}$

ν_2 \ ν_1	Degrees of freedom for the numerator (ν_1)																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	2.88	2.49	2.28	2.14	2.03	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

Degrees of freedom for the denominator (ν_2)

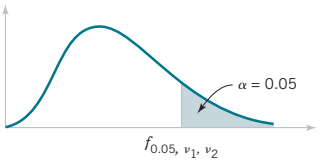


TABLE VI **Percentage Points f_{α,ν_1,ν_2} of the F Distribution (*continued*)**

f_{α,ν_1,ν_2}

ν_2 \ ν_1	Degrees of freedom for the numerator (ν_1)																		
Degrees of freedom for the denominator (ν_2)	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.55	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

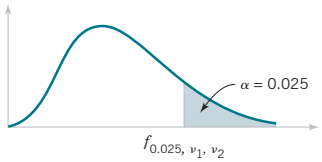


TABLE VI **Percentage Points f_{α,ν_1,ν_2} of the *F* Distribution (*continued*)**

f_{α,ν_1,ν_2}

ν_2 \ ν_1	Degrees of freedom for the numerator (ν_1)																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

Degrees of freedom for the denominator (ν_2)

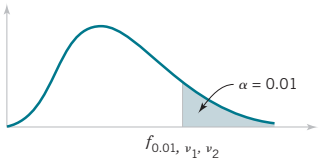
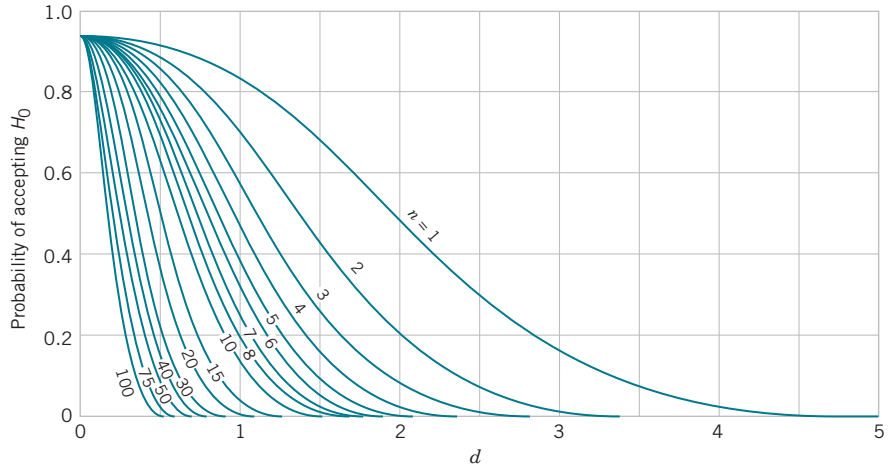


TABLE VI Percentage Points f_{α,ν_1,ν_2} of the F Distribution (*continued*)

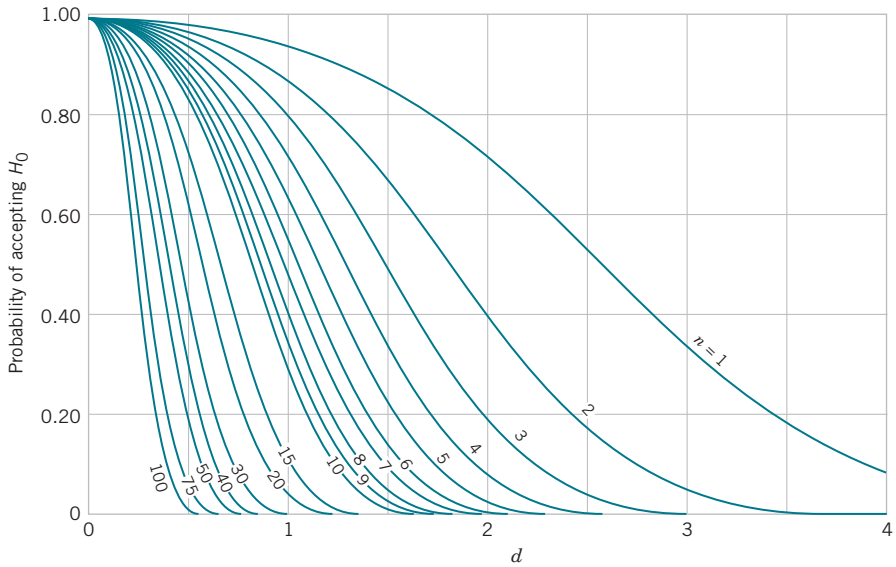
f_{α,ν_1,ν_2}

ν_2 \ ν_1	Degrees of freedom for the numerator (ν_1)																		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**∞**
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.00	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.46
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.36	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.59
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

Degrees of freedom for the denominator (ν_2)



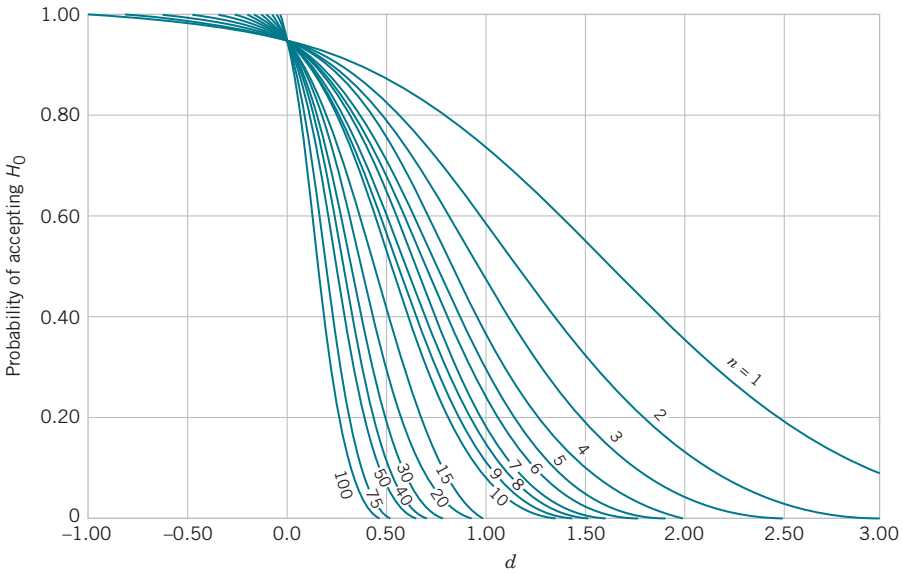
(*a*) O.C. curves for different values of n for the two-sided normal test for a level of significance $\alpha = 0.05$.



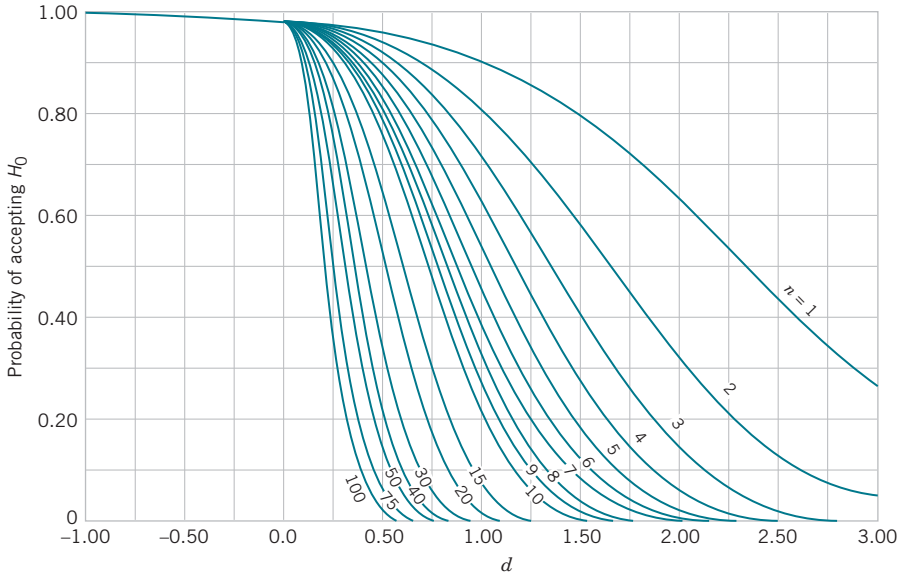
(*b*) O.C. curves for different values of n for the two-sided normal test for a level of significance $\alpha = 0.01$.

Source: Charts VI*a*, *e*, *f*, *k*, *m*, and *q* are reproduced with permission from "Operating Characteristics for the Common Statistical Tests of Significance," by C. L. Ferris, F. E. Grubbs, and C. L. Weaver, *Annals of Mathematical Statistics*, June 1946.
Charts VI*b*, *c*, *d*, *g*, *h*, *i*, *j*, *l*, *n*, *o*, *p*, and *r* are reproduced with permission from *Engineering Statistics*, 2nd Edition, by A. H. Bowker and G. J. Lieberman, Prentice-Hall, 1972.

CHART VII **Operating Characteristic Curves**

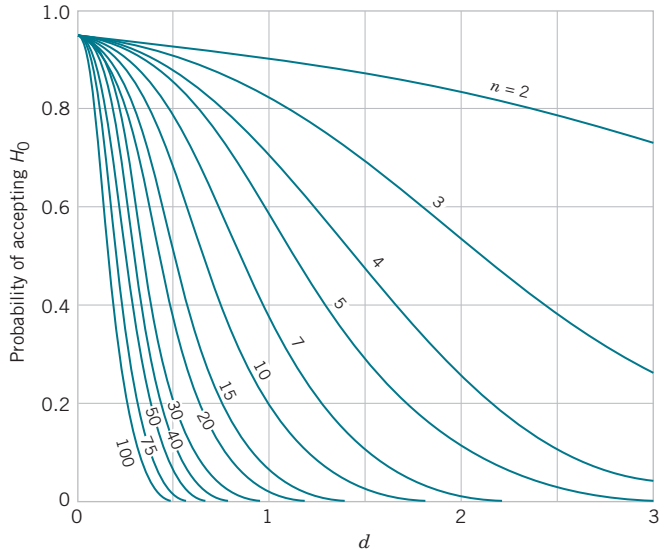


(*c*) O.C. curves for different values of n for the one-sided normal test for a level of significance $\alpha = 0.05$.

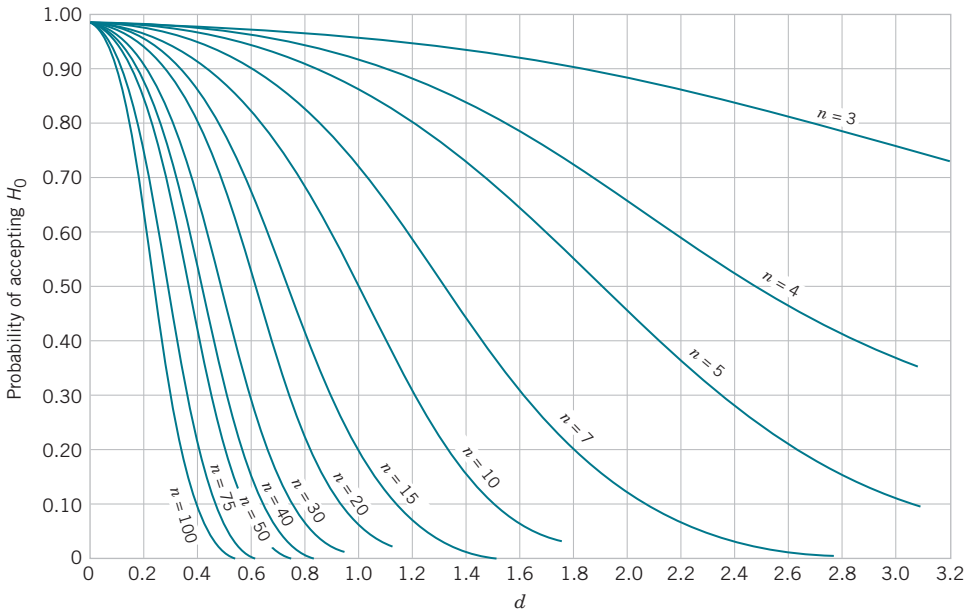


(*d*) O.C. curves for different values of n for the one-sided normal test for a level of significance $\alpha = 0.01$.

CHART VII **Operating Characteristic Curves (*continued*)**

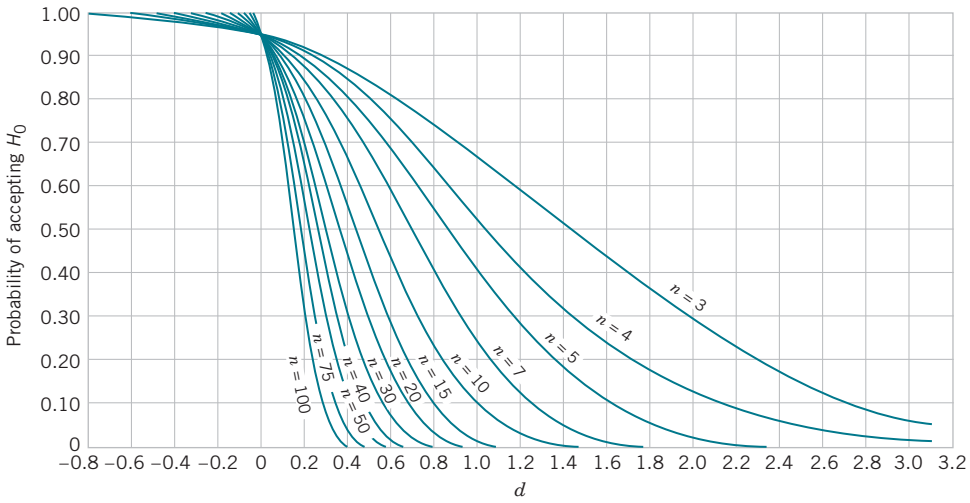


(*e*) O.C. curves for different values of n for the two-sided t-test for a level of significance $\alpha = 0.05$.

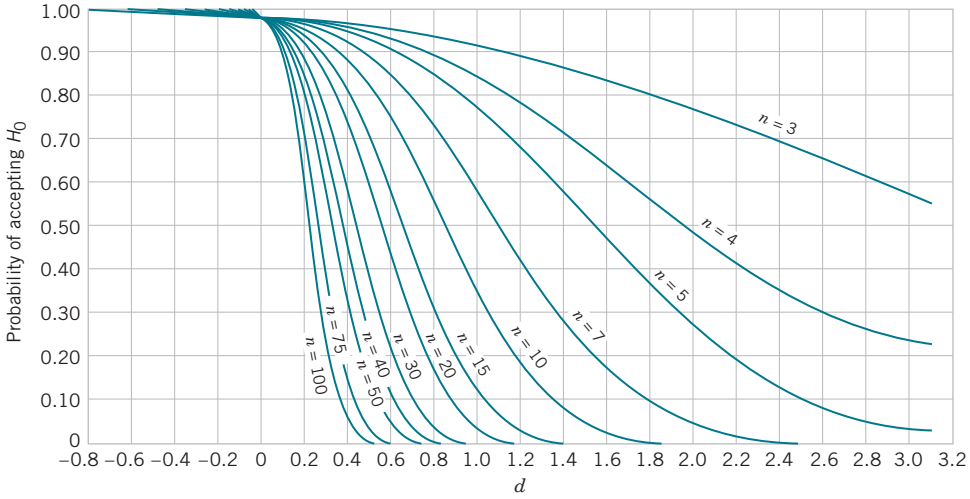


(*f*) O.C. curves for different values of n for the two-sided t-test for a level of significance $\alpha = 0.01$.

CHART VII **Operating Characteristic Curves (*continued*)**

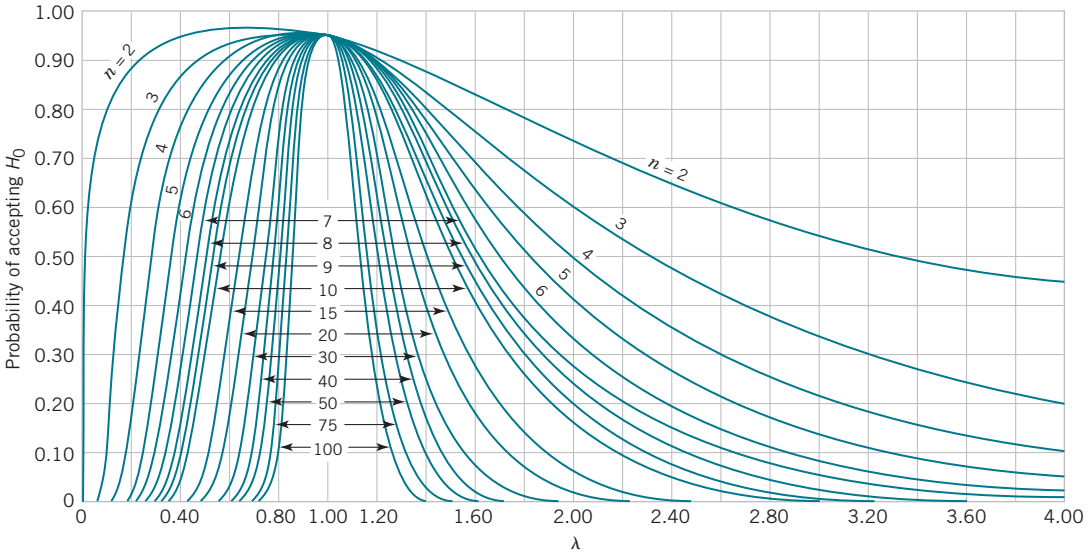


(*g*) O.C. curves for different values of n for the one-sided t-test for a level of significance $\alpha = 0.05$.

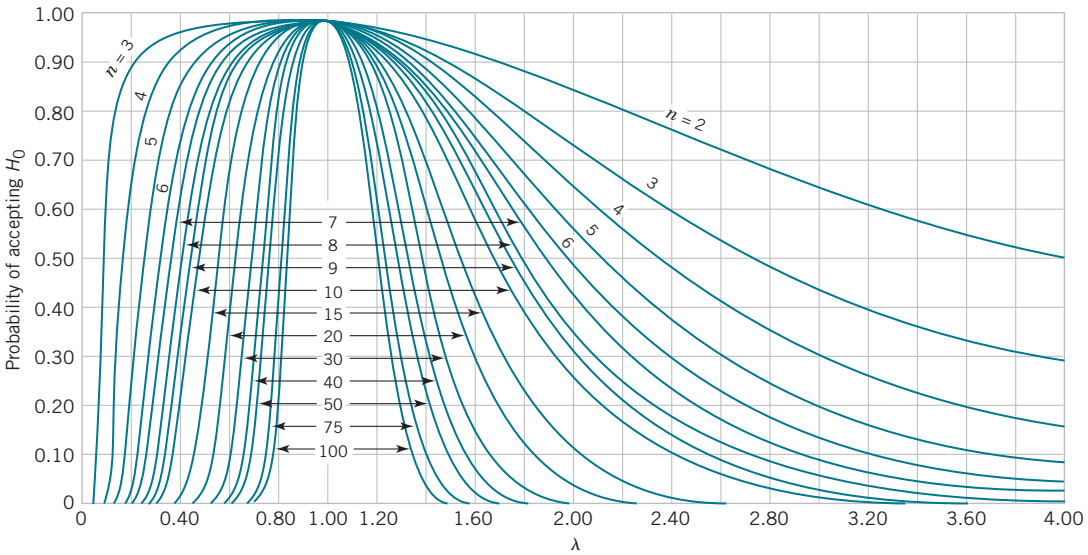


(*h*) O.C. curves for different values of n for the one-sided t-test for a level of significance $\alpha = 0.01$.

CHART VII **Operating Characteristic Curves (*continued*)**

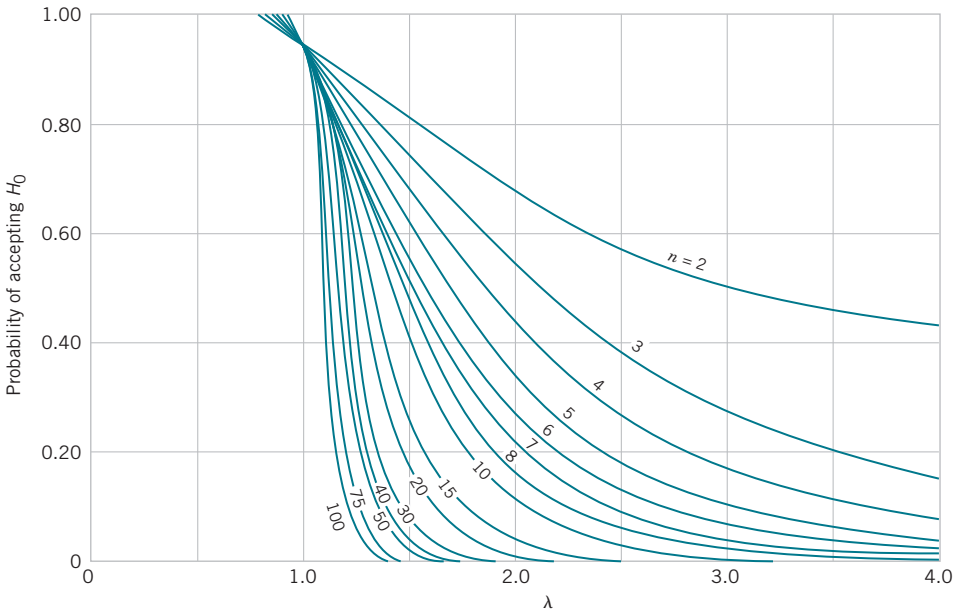


(*i*) O.C. curves for different values of n for the two-sided chi-square test for a level of significance $\alpha = 0.05$.

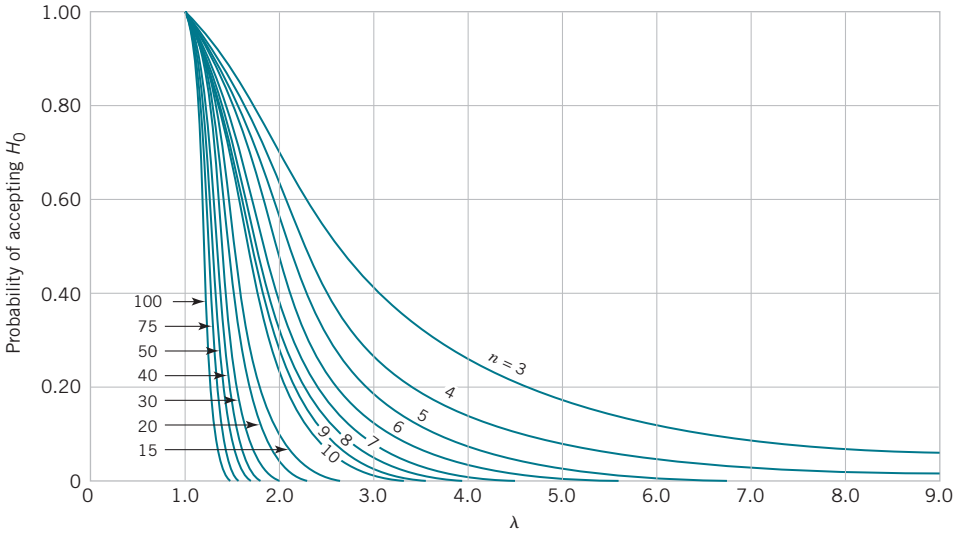


(*j*) O.C. curves for different values of n for the two-sided chi-square test for a level of significance $\alpha = 0.01$.

CHART VII **Operating Characteristic Curves (*continued*)**

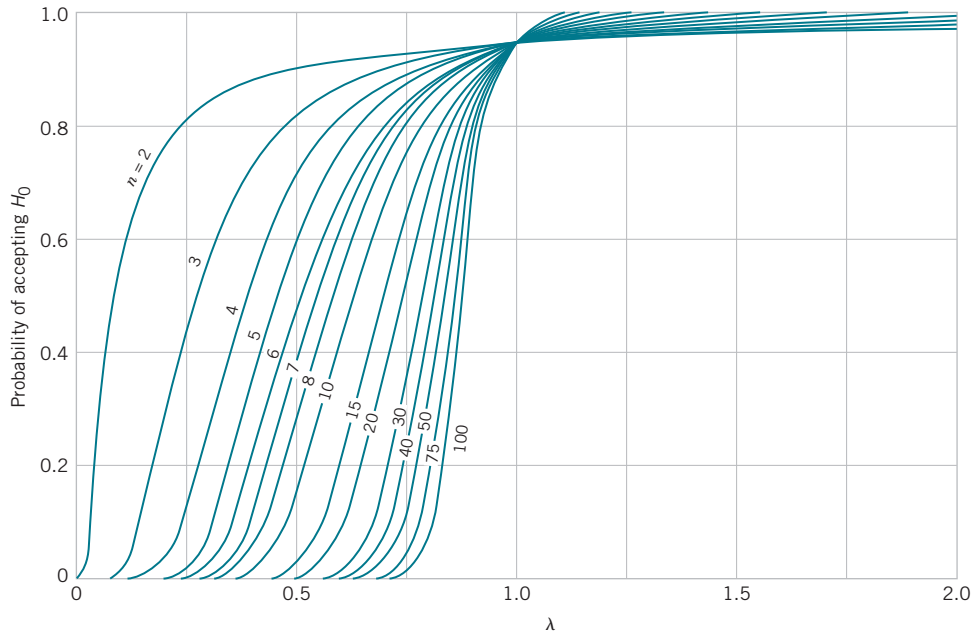


(*k*) O.C. curves for different values of *n* for the one-sided (upper-tail) chi-square test for a level of significance $\alpha = 0.05$.

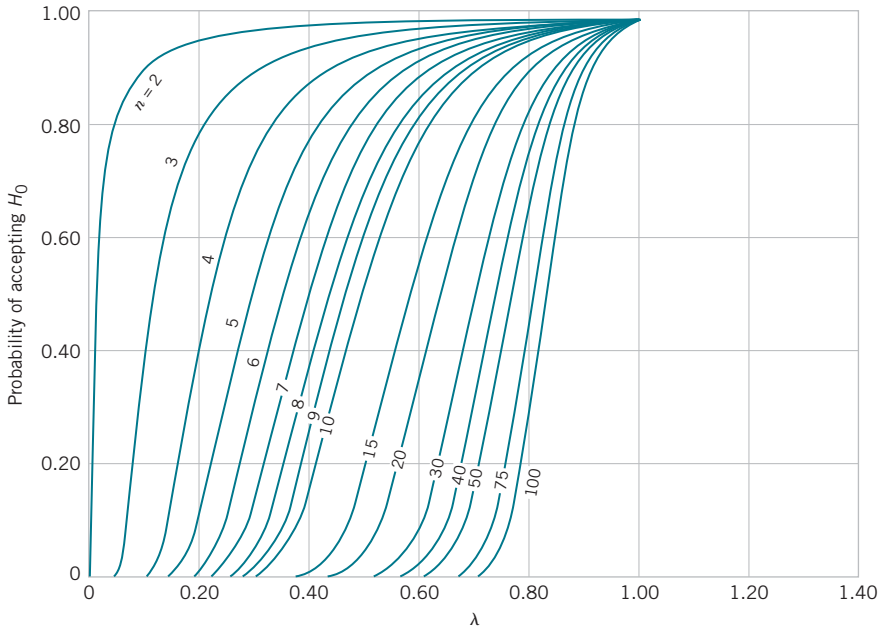


(*l*) O.C. curves for different values of *n* for the one-sided (upper-tail) chi-square test for a level of significance $\alpha = 0.01$.

CHART VII **Operating Characteristic Curves (*continued*)**

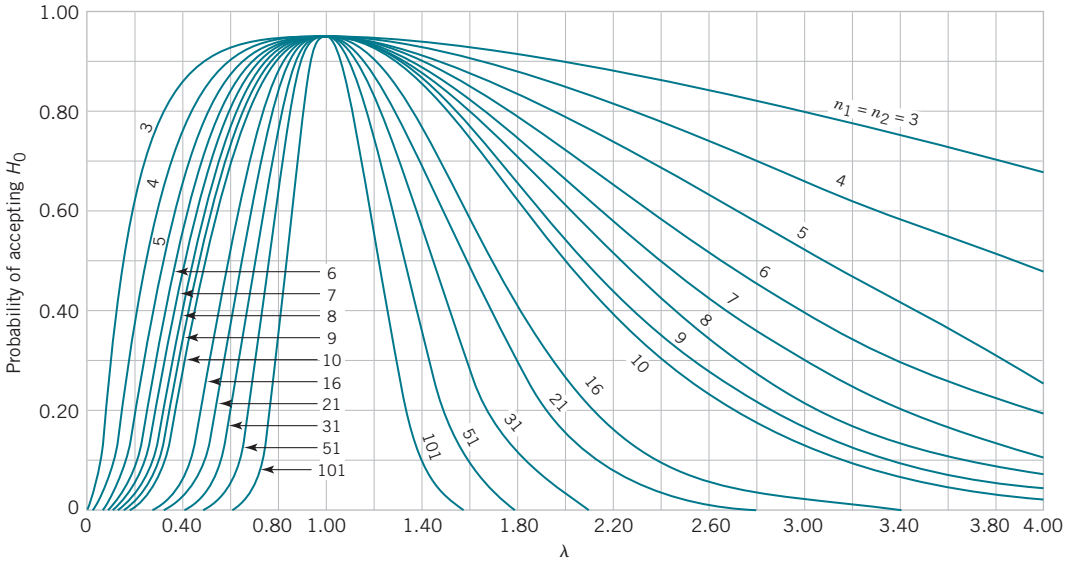


(*m*) O.C. curves for different values of *n* for the one-sided (lower-tail) chi-square test for a level of significance $\alpha = 0.05$.

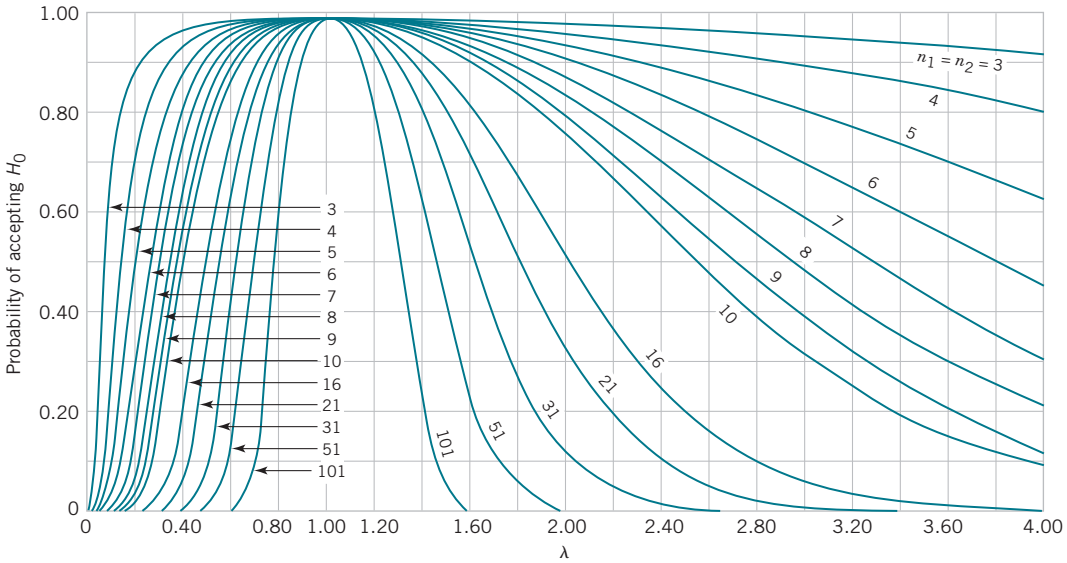


(*n*) O.C. curves for different values of *n* for the one-sided (lower-tail) chi-square test for a level of significance $\alpha = 0.01$.

CHART VII **Operating Characteristic Curves (*continued*)**

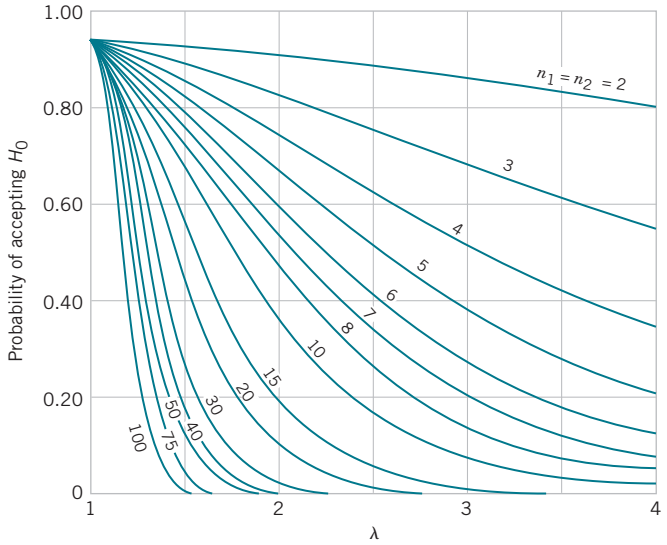


(*o*) O.C. curves for different values of n for the two-sided F-test for a level of significance $\alpha = 0.05$.

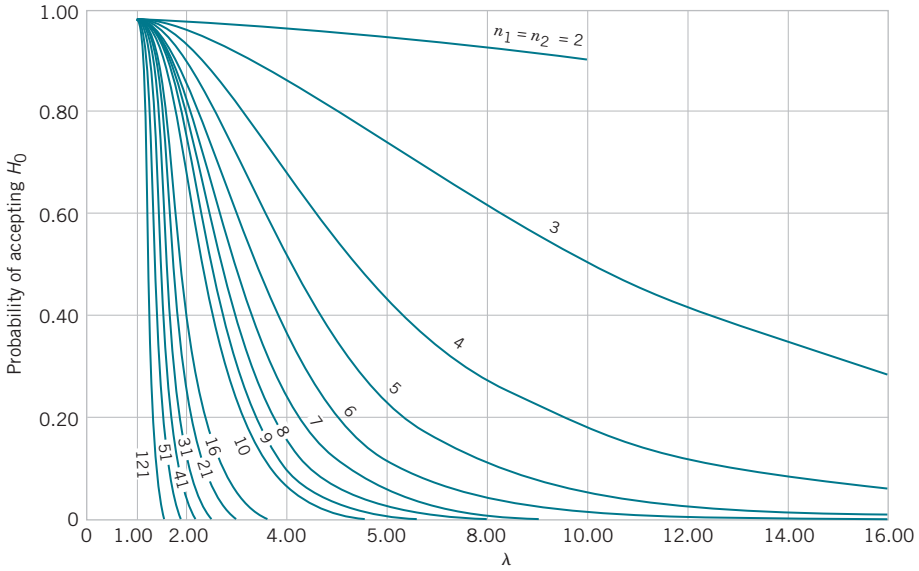


(*p*) O.C. curves for different values of n for the two-sided F-test for a level of significance $\alpha = 0.01$.

CHART VII **Operating Characteristic Curves (*continued*)**



(*q*) O.C. curves for different values of *n* for the one-sided *F*-test for a level of significance $\alpha = 0.05$.



(*r*) O.C. curves for different values of *n* for the one-sided *F*-test for a level of significance $\alpha = 0.01$.

CHART VII **Operating Characteristic Curves** (***continued***)

TABLE VIII Critical Values for the Sign Test

r_α^*

n \ α	0.10 0.05	0.05 0.025	0.01 0.005	Two-sided tests One-sided tests	n \ α	0.10 0.05	0.05 0.025	0.01 0.005	Two-sided tests One-sided tests
5	0				23	7	6	4	
6	0	0			24	7	6	5	
7	0	0			25	7	7	5	
8	1	0	0		26	8	7	6	
9	1	1	0		27	8	7	6	
10	1	1	0		28	9	8	6	
11	2	1	0		29	9	8	7	
12	2	2	1		30	10	9	7	
13	3	2	1		31	10	9	7	
14	3	2	1		32	10	9	8	
15	3	3	2		33	11	10	8	
16	4	3	2		34	11	10	9	
17	4	4	2		35	12	11	9	
18	5	4	3		36	12	11	9	
19	5	4	3		37	13	12	10	
20	5	5	3		38	13	12	10	
21	6	5	4		39	13	12	11	
22	6	5	4		40	14	13	11	

TABLE IX Critical Values for the Wilcoxon Signed-Rank Test

w_α^*

n* \ α	0.10 0.05	0.05 0.025	0.02 0.01	0.01 0.005	Two-sided tests One-sided tests
4					
5	0				
6	2	0			
7	3	2	0		
8	5	3	1	0	
9	8	5	3	1	
10	10	8	5	3	
11	13	10	7	5	
12	17	13	9	7	
13	21	17	12	9	
14	25	21	15	12	
15	30	25	19	15	
16	35	29	23	19	
17	41	34	27	23	
18	47	40	32	27	
19	53	46	37	32	
20	60	52	43	37	
21	67	58	49	42	
22	75	65	55	48	
23	83	73	62	54	
24	91	81	69	61	
25	100	89	76	68	

*If $n > 25$, W^- (or W^-) is approximately normally distributed with mean $n(n + 1)/4$ and variance $n(n + 1)(2n + 1)/24$.

TABLE X Critical Values for the Wilcoxon Rank-Sum Test

$w_{0.05}$

n_2 \ n_1*	4	5	6	7	8	9	10	11	12	13	14	15
4	10											
5	11	17										
6	12	18	26									
7	13	20	27	36								
8	14	21	29	38	49							
9	15	22	31	40	51	63						
10	15	23	32	42	53	65	78					
11	16	24	34	44	55	68	81	96				
12	17	26	35	46	58	71	85	99	115			
13	18	27	37	48	60	73	88	103	119	137		
14	19	28	38	50	63	76	91	106	123	141	160	
15	20	29	40	52	65	79	94	110	127	145	164	185
16	21	31	42	54	67	82	97	114	131	150	169	
17	21	32	43	56	70	84	100	117	135	154		
18	22	33	45	58	72	87	103	121	139			
19	23	34	46	60	74	90	107	124				
20	24	35	48	62	77	93	110					
21	25	37	50	64	79	95						
22	26	38	51	66	82							
23	27	39	53	68								
24	28	40	55									
25	28	42										
26	29											
27												
28												

*For n_1 and $n_2 > 8$, W_1 is approximately normally distributed with mean $\frac{1}{2}n_1\,(n_1 + n_2 + 1)$ and variance $n_1 n_2\,(n_1 + n_2 + 1)/12$.

TABLE X Critical Values for the Wilcoxon Rank-Sum Test (*continued*)

$w_{0.01}$

n_2 \ n_1*	4	5	6	7	8	9	10	11	12	13	14	15
5		15										
6	10	16	23									
7	10	17	24	32								
8	11	17	25	34	43							
9	11	18	26	35	45	56						
10	12	19	27	37	47	58	71					
11	12	20	28	38	49	61	74	87				
12	13	21	30	40	51	63	76	90	106			
13	14	22	31	41	53	65	79	93	109	125		
14	14	22	32	43	54	67	81	96	112	129	147	
15	15	23	33	44	56	70	84	99	115	133	151	171
16	15	24	34	46	58	72	86	102	119	137	155	
17	16	25	36	47	60	74	89	105	122	140		
18	16	26	37	49	62	76	92	108	125			
19	17	27	38	50	64	78	94	111				
20	18	28	39	52	66	81	97					
21	18	29	40	53	68	83						
22	19	29	42	55	70							
23	19	30	43	57								
24	20	31	44									
25	20	32										
26	21											
27												
28												

TABLE XI **Factors for Constructing Variables Control Charts**

Factor for Control Limits							
	$\bar{X}$ Chart			R Chart		S Chart	
n^*	A_1	A_2	d_2	D_3	D_4	C_4	n
2	3.760	1.880	1.128	0	3.267	0.7979	2
3	2.394	1.023	1.693	0	2.575	0.8862	3
4	1.880	.729	2.059	0	2.282	0.9213	4
5	1.596	.577	2.326	0	2.115	0.9400	5
6	1.410	.483	2.534	0	2.004	0.9515	6
7	1.277	.419	2.704	.076	1.924	0.9594	7
8	1.175	.373	2.847	.136	1.864	0.9650	8
9	1.094	.337	2.970	.184	1.816	0.9693	9
10	1.028	.308	3.078	.223	1.777	0.9727	10
11	.973	.285	3.173	.256	1.744	0.9754	11
12	.925	.266	3.258	.284	1.716	0.9776	12
13	.884	.249	3.336	.308	1.692	0.9794	13
14	.848	.235	3.407	.329	1.671	0.9810	14
15	.816	.223	3.472	.348	1.652	0.9823	15
16	.788	.212	3.532	.364	1.636	0.9835	16
17	.762	.203	3.588	.379	1.621	0.9845	17
18	.738	.194	3.640	.392	1.608	0.9854	18
19	.717	.187	3.689	.404	1.596	0.9862	19
20	.697	.180	3.735	.414	1.586	0.9869	20
21	.679	.173	3.778	.425	1.575	0.9876	21
22	.662	.167	3.819	.434	1.566	0.9882	22
23	.647	.162	3.858	.443	1.557	0.9887	23
24	.632	.157	3.895	.452	1.548	0.9892	24
25	.619	.153	3.931	.459	1.541	0.9896	25

*$n > 25 : A_1 = 3/\sqrt{n}$ where n = number of observations in sample.

TABLE XII **Factors for Tolerance Intervals**

Values of k for Two-Sided Intervals									
	Confidence Level								
	0.90			0.95			0.99		
	Probability of Coverage								
Sample Size	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
2	15.978	18.800	24.167	32.019	37.674	48.430	160.193	188.491	242.300
3	5.847	6.919	8.974	8.380	9.916	12.861	18.930	22.401	29.055
4	4.166	4.943	6.440	5.369	6.370	8.299	9.398	11.150	14.527
5	3.949	4.152	5.423	4.275	5.079	6.634	6.612	7.855	10.260
6	3.131	3.723	4.870	3.712	4.414	5.775	5.337	6.345	8.301
7	2.902	3.452	4.521	3.369	4.007	5.248	4.613	5.488	7.187
8	2.743	3.264	4.278	3.136	3.732	4.891	4.147	4.936	6.468
9	2.626	3.125	4.098	2.967	3.532	4.631	3.822	4.550	5.966
10	2.535	3.018	3.959	2.839	3.379	4.433	3.582	4.265	5.594
11	2.463	2.933	3.849	2.737	3.259	4.277	3.397	4.045	5.308
12	2.404	2.863	3.758	2.655	3.162	4.150	3.250	3.870	5.079
13	2.355	2.805	3.682	2.587	3.081	4.044	3.130	3.727	4.893
14	2.314	2.756	3.618	2.529	3.012	3.955	3.029	3.608	4.737
15	2.278	2.713	3.562	2.480	2.954	3.878	2.945	3.507	4.605
16	2.246	2.676	3.514	2.437	2.903	3.812	2.872	3.421	4.492
17	2.219	2.643	3.471	2.400	2.858	3.754	2.808	3.345	4.393
18	2.194	2.614	3.433	2.366	2.819	3.702	2.753	3.279	4.307
19	2.172	2.588	3.399	2.337	2.784	3.656	2.703	3.221	4.230
20	2.152	2.564	3.368	2.310	2.752	3.615	2.659	3.168	4.161
21	2.135	2.543	3.340	2.286	2.723	3.577	2.620	3.121	4.100
22	2.118	2.524	3.315	2.264	2.697	3.543	2.584	3.078	4.044
23	2.103	2.506	3.292	2.244	2.673	3.512	2.551	3.040	3.993
24	2.089	2.489	3.270	2.225	2.651	3.483	2.522	3.004	3.947
25	2.077	2.474	3.251	2.208	2.631	3.457	2.494	2.972	3.904
30	2.025	2.413	3.170	2.140	2.529	3.350	2.385	2.841	3.733
40	1.959	2.334	3.066	2.052	2.445	3.213	2.247	2.677	3.518
50	1.916	2.284	3.001	1.996	2.379	3.126	2.162	2.576	3.385
60	1.887	2.248	2.955	1.958	2.333	3.066	2.103	2.506	3.293
70	1.865	2.222	2.920	1.929	2.299	3.021	2.060	2.454	3.225
80	1.848	2.202	2.894	1.907	2.272	2.986	2.026	2.414	3.173
90	1.834	2.185	2.872	1.889	2.251	2.958	1.999	2.382	3.130
100	1.822	2.172	2.854	1.874	2.233	2.934	1.977	2.355	3.096

TABLE XII Factors for Tolerance Intervals (*continued*)

Values of k for Two-Sided Intervals									
	Confidence Level								
	0.90			0.95			0.99		
	Probability of Coverage								
Sample Size	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
2	10.253	13.090	18.500	20.581	26.260	37.094	103.029	131.426	185.617
3	4.258	5.311	7.340	6.155	7.656	10.553	13.995	17.370	23.896
4	3.188	3.957	5.438	4.162	5.144	7.042	7.380	9.083	12.387
5	2.742	3.400	4.666	3.407	4.203	5.741	5.362	6.578	8.939
6	2.494	3.092	4.243	3.006	3.708	5.062	4.411	5.406	7.335
7	2.333	2.894	3.972	2.755	3.399	4.642	3.859	4.728	6.412
8	2.219	2.754	3.783	2.582	3.187	4.354	3.497	4.285	5.812
9	2.133	2.650	3.641	2.454	3.031	4.143	3.240	3.972	5.389
10	2.066	2.568	3.532	2.355	2.911	3.981	3.048	3.738	5.074
11	2.011	2.503	3.443	2.275	2.815	3.852	2.898	3.556	4.829
12	1.966	2.448	3.371	2.210	2.736	3.747	2.777	3.410	4.633
13	1.928	2.402	3.309	2.155	2.671	3.659	2.677	3.290	4.472
14	1.895	2.363	3.257	2.109	2.614	3.585	2.593	3.189	4.337
15	1.867	2.329	3.212	2.068	2.566	3.520	2.521	3.102	4.222
16	1.842	2.299	3.172	2.033	2.524	3.464	2.459	3.028	4.123
17	1.819	2.272	3.137	2.002	2.486	3.414	2.405	2.963	4.037
18	1.800	2.249	3.105	1.974	2.453	3.370	2.357	2.905	3.960
19	1.782	2.227	3.077	1.949	2.423	3.331	2.314	2.854	3.892
20	1.765	2.028	3.052	1.926	2.396	3.295	2.276	2.808	3.832
21	1.750	2.190	3.028	1.905	2.371	3.263	2.241	2.766	3.777
22	1.737	2.174	3.007	1.886	2.349	3.233	2.209	2.729	3.727
23	1.724	2.159	2.987	1.869	2.328	3.206	2.180	2.694	3.681
24	1.712	2.145	2.969	1.853	2.309	3.181	2.154	2.662	3.640
25	1.702	2.132	2.952	1.838	2.292	3.158	2.129	2.633	3.601
30	1.657	2.080	2.884	1.777	2.220	3.064	2.030	2.515	3.447
40	1.598	2.010	2.793	1.697	2.125	2.941	1.902	2.364	3.249
50	1.559	1.965	2.735	1.646	2.065	2.862	1.821	2.269	3.125
60	1.532	1.933	2.694	1.609	2.022	2.807	1.764	2.202	3.038
70	1.511	1.909	2.662	1.581	1.990	2.765	1.722	2.153	2.974
80	1.495	1.890	2.638	1.559	1.964	2.733	1.688	2.114	2.924
90	1.481	1.874	2.618	1.542	1.944	2.706	1.661	2.082	2.883
100	1.470	1.861	2.601	1.527	1.927	2.684	1.639	2.056	2.850

Appendix B

Bibliography

Introductory Works and Graphical Methods

Chambers, J., Cleveland, W., Kleiner, B., and P. Tukey (1983), *Graphical Methods for Data Analysis*, Wadsworth & Brooks/Cole, Pacific Grove, CA. A very well-written presentation of graphical methods in statistics.

Freedman, D., Pisani, R., Purves R., and A. Adbikari (2007), *Statistics*, 4th ed., Norton, New York. An excellent introduction to statistical thinking, requiring minimal mathematical background.

Hoaglin, D., Mosteller, F., and J. Tukey (1983), *Understanding Robust and Exploratory Data Analysis*, John Wiley & Sons, New York. Good discussion and illustration of techniques such as stem-and-leaf displays and box plots.

Tanur, J., et al. (eds.) (1989), *Statistics: A Guide to the Unknown*, 3rd edition, Wadsworth & Brooks/Cole, Pacific Grove, CA. Contains a collection of short nonmathematical articles describing different applications of statistics.

Tukey, J. (1977), *Exploratory Data Analysis*, Addison-Wesley, Reading, MA. Introduces many new descriptive and analytical methods. Not extremely easy to read.

Probability

Hoel, P. G., Port, S. C., and C. J. Stone (1971), *Introduction to Probability Theory*, Houghton Mifflin, Boston. A well-written and comprehensive treatment of probability theory and the standard discrete and continuous distributions.

Olkin, I., Derman, C., and L. Gleser (1994), *Probability Models and Applications*, 2nd ed., Macmillan, New York. A comprehensive treatment of probability at a higher mathematical level than this book.

Mosteller, F., Rourke, R., and G. Thomas (1970), *Probability with Statistical Applications*, 2nd ed., Addison-Wesley, Reading, MA. A precalculus introduction to probability with many excellent examples.

Ross, S. (2009), *A First Course in Probability*, 8th ed., Prentice-Hall, Englewood Cliffs, NJ. More mathematically sophisticated than this book, but has many excellent examples and exercises.

Mathematical Statistics

Efron, B., and R. Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman and Hall, New York. An important reference on this useful but computer-intensive technique.

Hoel, P. G. (1984), *Introduction to Mathematical Statistics*, 5th ed., John Wiley & Sons, New York. An outstanding introductory book, well written, and generally easy to understand.

Hogg, R., and A. Craig (2004), *Introduction to Mathematical Statistics*, 6th ed., Prentice-Hall, Englewood Cliffs, NJ. Another classic work on the mathematical principles of statistics; higher level than the Hoel book, but contains excellent discussions of estimation and hypothesis testing.

Larsen, R., and M. Marx (1986), *Introduction to Mathematical Statistics*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ. Written at a relatively low mathematical level, very readable.

Larson, H. J. (1982), *Introduction to Probability Theory and Statistical Inference*, 3rd ed., John Wiley & Sons, New York. An extremely well-written book that gives broad coverage to many aspects of probability and mathematical statistics.

Engineering Statistics

Devore, J. L. (2008), *Probability and Statistics for Engineering and the Sciences*, 8th ed., Duxbury & Brooks/Cole, Pacific Grove, CA. Covers many of the same topics as this text, but at a slightly higher mathematical level. Many of the examples and exercises involve applications to biological and life sciences.

Hines, W. W., and D. C. Montgomery (1990), *Probability and Statistics in Engineering and Management Science*, 3rd ed., John Wiley & Sons, New York. Covers many of the same topics as this book. More emphasis on probability and a higher mathematical level.

Ross, S. (1987), *Introduction to Probability and Statistics for Engineers and Scientists*, John Wiley & Sons, New York. More tightly written and mathematically oriented than this book, but contains some good examples.

Walpole, R. E., Myers, R. H., and S. L. Myers (2002), *Probability and Statistics for Engineers and Scientists*, 7th ed., Prentice-Hall, Inc., Upper Saddle River, New Jersey. A very well-written book at about the same level as this one.

Regression Analysis

Daniel, C., and F. Wood (1980), *Fitting Equations to Data*, 2nd ed., John Wiley & Sons, New York. An excellent reference containing many insights on data analysis.

Draper, N., and H. Smith (1998), *Applied Regression Analysis*, 3rd ed., John Wiley & Sons, New York. A comprehensive book on regression written for statistically oriented readers.

Kutner, M.H., Nachtsheim, C.J., Neter, J., and W. Li (2005), *Applied Linear Statistical Models*, 5th ed., McGraw-Hill/Irwin, Columbus, OH. The first part of the book is an introduction to simple and multiple linear regression. The orientation is to business and economics.

Montgomery, D. C., Peck, E. A., and G. G. Vining (2012), *Introduction to Linear Regression Analysis*, 5th ed., John Wiley & Sons, New York. A comprehensive book on regression written for engineers and physical scientists.

Myers, R. H. (1990), *Classical and Modern Regression with Applications*, 2nd ed., PWS-Kent, Boston. Contains many examples with annotated SAS output. Very well written.

Design of Experiments

Box, G. E. P., Hunter, W. G., and J. S. Hunter (2005), *Statistics for Experimenters*, 2nd ed., John Wiley & Sons, New York. An excellent introduction to the subject for those readers desiring a statistically oriented treatment. Contains many useful suggestions for data analysis.

Mason, R. L., Gunst, R. F., and J. F. Hess (2003), *Statistical Design and Analysis of Experiments*, 2nd ed., John Wiley & Sons, New York. A comprehensive book covering basic statistics, hypothesis testing and confidence intervals, elementary aspects of experimental design, and regression analysis.

Montgomery, D. C. (2017), *Design and Analysis of Experiments*, 9th ed., John Wiley & Sons, New York. Written at the same level as the Box, Hunter, and Hunter book, but focused on engineering applications.

Nonparametric Statistics

Conover, W. J. (1998), *Practical Nonparametric Statistics*, 3rd ed., John Wiley & Sons, New York. An excellent exposition of the methods of nonparametric statistics; many good examples and exercises.

Hollander, M., and D. Wolfe (1999), *Nonparametric Statistical Methods*, 2nd ed., John Wiley & Sons, New York. A good reference book, with a very useful set of tables.

Statistical Quality Control and Related Methods

Duncan, A. J. (1986), *Quality Control and Industrial Statistics*, 5th ed., Richard D. Irwin, Homewood, Illinois. A classic book on the subject.

Grant, E. L., and R. S. Leavenworth (1996), *Statistical Quality Control*, 7th ed., McGraw-Hill, New York. One of the first books on the subject; contains many good examples.

John, P. W. M. (1990), *Statistical Methods in Engineering and Quality Improvement*, John Wiley & Sons, New York. Not a methods book, but a well-written presentation of statistical methodology for quality improvement.

Montgomery, D. C. (2013), *Introduction to Statistical Quality Control*, 7th ed., John Wiley & Sons, New York. A modern comprehensive treatment of the subject written at the same level as this book.

Nelson, W. (2003), *Applied Life Data Analysis*, John Wiley & Sons, New York. Contains many examples of using statistical methods for the study of failure data; a good reference for the statistical aspects of reliability engineering and the special probability distributions used in that field.

Ryan, T. P. (2000), *Statistical Methods for Quality Improvement*, 2nd ed., John Wiley & Sons, New York. Gives broad coverage of the field, with some emphasis on newer techniques.

Wadsworth, H. M., Stephens, K. S., and A. B. Godfrey (2001), *Modern Methods for Quality Control and Improvement*, 2nd ed., John Wiley & Sons, New York. A comprehensive treatment of statistical methods for quality improvement at a somewhat higher level than this book.

Western Electric Company (1956), *Statistical Quality Control Handbook*, Western Electric Company, Inc., Indianapolis, Indiana. An oldie but a goodie.

Glossary

100(1 − α)% confidence interval on μ A statement that contains μ that will be true 100(1 − α)% of the time that the statement is made.

100(1 − α)% confidence interval on σ^2 A statement that contains σ^2 that will be true 100(1 − α)% of the time that the statement is made.

3-sigma control limit Control limits on a Shewhart control chart located a distance of plus or minus 3 standard deviations of the statistic.

Acceptance region In hypothesis testing, a region in the sample space of the test statistic such that if the test statistic falls within it, the null hypothesis cannot be rejected. This terminology is used because rejection of H_0 is always a strong conclusion and acceptance of H_0 is generally a weak conclusion.

Addition rule A formula used to determine the probability of the union of two (or more) events from the probabilities of the events and their intersection(s).

Adjusted R^2 A variation of the R^2 statistic that compensates for the number of parameters in a regression model. Essentially, the adjustment is a penalty for increasing the number of parameters in the model.

Agresti-Coull CI Alternate way to construct a confidence interval on a binomial proportion from the traditional approach, proposed by Agresti and Coull in 1998.

Alias In a fractional factorial experiment when certain factor effects cannot be estimated uniquely, they are said to be aliased.

All possible (subsets) regressions A method of variable selection in regression that examines all possible subsets of the candidate regressor variables. Efficient computer algorithms have been developed for implementing all possible regressions.

Alternate fraction The fraction with the minus sign in a defining relation.

Alternative hypothesis In statistical hypothesis testing, this is a hypothesis other than the one that is being tested. The alternative hypothesis contains feasible conditions, whereas the null hypothesis specifies conditions that are under test.

Analysis of variance (ANOVA) A method of decomposing the total variability in a set of observations, as measured by the sum of the squares of these observations from their average, into component sums of squares that are associated with specific defined sources of variation.

Analytic study A study in which a sample from a population is used to make inference to a future population. Stability needs to be assumed. *See* Enumerative study.

Assignable cause The portion of the variability in a set of observations that can be traced to specific causes, such as operators, materials, or equipment. Also called a special cause.

Attribute control chart Any control chart for a discrete random variable. *See* Variables control chart.

Average run length, or ARL The average number of samples taken in a process monitoring or inspection scheme until the scheme signals that the process is operating at a level different from the level in which it began.

Axioms of probability A set of rules that probabilities defined on a sample space must follow. *See* Probability.

Backward elimination A method of variable selection in regression that begins with all of the candidate regressor variables in the model and eliminates the insignificant regressors one at a time until only significant regressors remain.

Bayes' estimator An estimator for a parameter obtained from a Bayesian method that uses a prior distribution for the parameter along with the conditional distribution of the data given the parameter to obtain the posterior distribution of the parameter. The estimator is obtained from the posterior distribution.

Bayes' theorem An equation for a conditional probability such as $P(A|B)$ in terms of the reverse conditional probability $P(B|A)$.

Bernoulli trials Sequences of independent trials with only two outcomes, generally called "success" and "failure," in which the probability of success remains constant.

Beta random variable The random variable X with a probability density function with parameters $\alpha > 0$ and $\beta > 0$.

Bias An effect that systematically distorts a statistical result or estimate, preventing it from representing the true quantity of interest.

Biased estimator *See* Unbiased estimator.

Binomial distribution A random experiment that consists of n Bernoulli trials such that the trials are independent, each trial results in only two possible outcomes, labeled as "success" and "failure"; the probability of a success in each trial, denoed as p, remains constant.

Binomial random variable A discrete random variable that equals the number of successes in a fixed number of Bernoulli trials.

Bivariate distribution The joint probability distribution of two random variables.

Bivariate normal distribution The joint distribution of two normal random variables.

Block In experimental design, a group of experimental units or material that is relatively homogeneous. The purpose of dividing experimental units into blocks is to produce an experimental design wherein variability within blocks is smaller than variability between blocks. This allows the factors of interest to be compared in an environment that has less variability than in an unblocked experiment.

Bootstrap Computer-intensive technique that treats data samples as a population.

Bootstrap estimate Estimate calculated from bootstrap samples.

Bootstrap samples Samples generated randomly from the probability distribution by computer.

Bootstrap standard error Sample standard deviation of the bootstrap estimate.

Box plot (or box and whisker plot) A graphical display of data in which the box contains the middle 50% of the data (the interquartile range) with the median dividing it, and the whiskers extend to the smallest and largest values (or some defined lower and upper limits).

***C* chart** An attribute control chart that plots the total number of defects per unit in a subgroup. Similar to a defects-per-unit or *U* chart.

Categorical data Data consisting of counts or observations that can be classified into categories. The categories may be descriptive.

Cause-and-effect Relationship with a clear causal link between an event and a result.

Cause-and-effect diagram A chart used to organize the various potential causes of a problem. Also called a fishbone diagram.

Center line A horizontal line on a control chart at the value that estimates the mean of the statistic plotted on the chart. *See* Control chart.

Center points Runs at the geometric center of a design.

Central composite design (CCD) A second-order response surface design in k variables consisting of a two-level factorial, $2k$ axial runs, and one or more center points. The two-level factorial portion of a CCD can be a fractional factorial design when k is large. The CCD is the most widely used design for fitting a second-order model.

Central limit theorem The simplest form of the central limit theorem states that the sum of **n** independently distributed random variables will tend to be normally distributed as **n** becomes large. It is a necessary and sufficient condition that none of the variances of the individual random variables are large in comparison to their sum. There are more general forms of the central theorem that allow infinite variances and correlated random variables, and there is a multivariate version of the theorem.

Chance cause The portion of the variability in a set of observations that is due to only random forces and which cannot be traced to specific sources, such as operators, materials, or equipment. Also called a *common cause*.

Chance causes Inherent or natural variability.

Chi-square test Any test of significance based on the chi-square distribution. The most common chi-square tests are (1) testing hypotheses about the variance or standard deviation of a normal distribution and (2) testing goodness of fit of a theoretical distribution to sample data.

Chi-squared distribution A special case of the gamma distribution in which $\lambda = 1/2$ and r equals one of the values 1/2, 1. 3/2, 2 …

Coefficient of determination *See* R^2.

Coefficient of multiple determination *See* Coefficient of determination.

Combination A subset selected without replacement from a set used to determine the number of outcomes in events and sample spaces.

Comparative experiment An experiment in which the treatments (experimental conditions) that are to be studied are included in the experiment. The data from the experiment are used to evaluate the treatments.

Completely randomized design (or experiment) A type of experimental design in which the treatments or design factors are assigned to the experimental units in a random manner. In designed experiments, a completely randomized design results from running all of the treatment combinations in random order.

Components of variance The individual components of the total variance that are attributable to specific sources. This usually refers to the individual variance components arising from a random or mixed model analysis of variance.

Conditional mean The mean of the conditional probability distribution of a random variable.

Conditional probability The probability of an event given that the random experiment produces an outcome in another event.

Conditional probability density function The probability density function of the conditional probability distribution of a continuous random variable.

Conditional probability mass function The probability mass function of the conditional probability distribution of a discrete random variable.

Conditional variance The variance of the conditional probability distribution of a random variable.

Confidence *See* Confidence level.

Confidence coefficient The probability $1 - \alpha$ associated with a confidence interval expressing the probability that the stated interval will contain the true parameter value.

Confidence interval If it is possible to write a probability statement of the form $P(L \leq \theta \leq U) = 1 - \alpha$ where L and U are functions of only the sample data and θ is a parameter, then the interval between L and U is called a confidence interval (or a $100(1 - \alpha)\%$ confidence interval). The interpretation is that a statement that the parameter θ lies in this interval will be true $100(1 - \alpha)\%$ of the times that such a statement is made.

Confidence limits (bounds), lower and upper End-points of a confidence interval.

Confounded with blocks *See* Confounding.

Confounding When a factorial experiment is run in blocks and the blocks are too small to contain a complete replicate of the experiment, one can run a fraction of the replicate in each block, but this results in losing information on some effects. These effects are linked with or confounded with the blocks. In general, when two factors are varied such that their individual effects cannot be determined separately, their effects are said to be confounded.

Contingency table A tabular arrangement expressing the assignment of members of a data set according to two or more categories or classification criteria.

Continuity correction A correction factor used to improve the approximation to binomial probabilities from a normal distribution.

Continuous random variable A random variable with an interval (either finite or infinite) of real numbers for its range.

Continuous uniform random variable A continuous random variable with range of a finite interval and a constant probability density function.

Contrast A linear function of treatment means with coefficients that total zero. A contrast is a summary of treatment means that is of interest in an experiment.

Control chart A graphical display used to monitor a process. It usually consists of a horizontal center line corresponding to the in-control value of the parameter that is being monitored and lower and upper control limits. The control limits are determined by statistical criteria and are not arbitrary, nor are they related to specification limits. If sample points fall within the control limits, the process is said to be in-control, or free from assignable causes. Points beyond the control limits indicate an out-of-control process; that is, assignable causes are likely present. This signals the need to find and remove the assignable causes.

Control limits *See* Control chart.

Controllable variables Variables in a process or experiment that can be controlled.

Cook's distance In regression, Cook's distance is a measure of the influence of each individual observation on the estimates of the regression model parameters. It expresses the distance that the vector of model parameter estimates with the ith observation removed lies from the vector of model parameter estimates based on all observations. Large values of Cook's distance indicate that the observation is influential.

Correlation In the most general usage, a measure of the interdependence among data. The concept may include more than two variables. The term is most commonly used in a narrow sense to express the relationship between quantitative variables or ranks.

Correlation coefficient A dimensionless measure of the linear association between two variables, usually lying in the interval from -1 to $+1$, with zero indicating the absence of correlation (but not necessarily the independence of the two variables).

Counting techniques Formulas used to determine the number of elements in sample spaces and events.

Covariance A measure of association between two random variables obtained as the expected value of the product of the two random variables around their means; that is, $\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$.

C_p statistic A measure of the total mean square error for the regression model.

Critical region In hypothesis testing, this is the portion of the sample space of a test statistic that will lead to rejection of the null hypothesis.

Critical value(s) The value of a statistic corresponding to a stated significance level as determined from the sampling distribution. For example, if $P(Z \geq z_{0.05}) = P(Z \geq 1.96) = 0.05$, then $z_{0.05} = 1.96$ is the critical value of z at the 0.025 level of significance.

Cumulative distribution function For a random variable X, the function of X defined as $P(X \leq x)$ that is used to specify the probability distribution.

Cumulative sum control chart (CUSUM) A control chart in which the point plotted at time t is the sum of the measured deviations from target for all statistics up to time t.

Decision theory Study of mathematical models for decision making.

Defect concentration diagram A quality tool that graphically shows the location of defects on a part or in a process.

Defects-per-unit control chart *See U* chart.

Defining contrast *See* Contrast.

Defining relation A subset of effects in a fractional factorial design that define the aliases in the design.

Degrees of freedom The number of independent comparisons that can be made among the elements of a sample. The term is analogous to the number of degrees of freedom for an object in a dynamic system, which is the number of independent coordinates required to determine the motion of the object.

Deming's 14 points A management philosophy promoted by W. Edwards Deming that emphasizes the importance of change and quality.

Design matrix A matrix that provides the tests that are to be conducted in an experiment.

Designed experiment An experiment in which the tests are planned in advance and the plans usually incorporate statistical models. *See* Experiment.

Discrete random variable A random variable with a finite (or countably infinite) range.

Discrete uniform distribution Where a random variable X, if each of the n values in its range, $x_1, x_2, \ldots x_n$, has equal probability.

Distance measure, Cook's distance measure Diagnostic equation for detecting influential observations.

Empirical model A model to relate a response to one or more regressors or factors that is developed from data obtained from the system.

Enumerative study A study in which a sample from a population is used to make inference to the population. *See* Analytic study.

Equivalence band Practical threshold or limit within which the mean performance is considered to be the same as the standard.

Equivalence interval Interval plus or minus δ.

Equivalence test Hypothesis test using two sets of one-sided alternative hypotheses to test equivalence.

Erlang random variable A continuous random variable that is the sum of a fixed number of independent, exponential random variables.

Error propagation An analysis of how the variance of the random variable that represents that output of a system depends on the variances of the inputs. A formula exists when the output is a linear function of the inputs and the formula is simplified if the inputs are assumed to be independent.

Error sum of squares In analysis of variance, this is the portion of total variability that is due to the random component in the data. It is usually based on replication of observations at certain treatment combinations in the experiment. It is sometimes called the residual sum of squares, although this is really a better term to use only when the sum of squares is based on the remnants of a model-fitting process and not on replication.

Estimated standard error Standard error involvng unknown parameters.

Event A subset of a sample space.

Expected cost criterion Criterion selects the action to minimize the expected cost.

Expected value The expected value of a random variable X is its long-term average or mean value. In the continuous case, the expected value of X is $E(X) = \int_{-\infty}^{\infty} xf(x)dx$ where $f(x)$ is the density function of the random variable X.

Exponential random variable A continuous random variable that is the time between events in a Poisson process.

Exponentially-weighted moving average (EWMA) Equation to calculate weighting factors which decrease exponentially.

Extra sum of squares method A method used in regression analysis to conduct a hypothesis test for the additional contribution of one or more variables to a model.

Factorial experiment A type of experimental design in which every level of one factor is tested in combination with every level of another factor. In general, in a factorial experiment, all possible combinations of factor levels are tested.

False alarm A signal from a control chart when no assignable causes are present.

Finite population correction factor A term in the formula for the variance of a hypergeometric random variable.

Fisher's least significant difference (LSD) method A series of pair-wise hypothesis tests of treatment means in an experiment to determine which means differ.

Fixed significance level Approach to conducting a single-sample t-test where the value of the test statistic falls in a critical region.

Fixed-effects model *See* Fixed factor.

Fixed-significance-level test Hypothesis test with a set criterion for an alternative hypothesis.

Forward selection A method of variable selection in regression, where variables are inserted one at a time into the model until no other variables that contribute significantly to the model can be found.

Fraction defective control chart *See P* chart.

Fractional factorial design *See* Fractional factorial experiment.

Fractional factorial experiment A type of factorial experiment in which not all possible treatment combinations are run. This is usually done to reduce the size of an experiment with several factors.

Frequency distribution An arrangement of the frequencies of observations in a sample or population according to the values that the observations take on.

Gamma function A function used in the probability density function of a gamma random variable that can be considered to extend factorials.

Gamma random variable A random variable that generalizes an Erlang random variable to noninteger values of the parameter r.

Gaussian distribution Another name for the normal distribution, based on the strong connection of Karl F. Gauss to the normal distribution; often used in physics and electrical engineering applications.

General regression significance test *See* Extra sum of squares method.

Generator Effects in a fractional factorial experiment that are used to construct the

experimental tests used in the experiment. The generators also define the aliases.

Geometric distribution In a series of Bernoulli trials (independent trials with constant probability p of a success), the random variable X that equals the number of trials until the first success is a geometric random variable.

Geometric random variable A discrete random variable that is the number of Bernoulli trials until a success occurs.

Goodness of fit In general, the agreement of a set of observed values and a set of theoretical values that depend on some hypothesis. The term is often used in fitting a theoretical distribution to a set of observations.

Grand mean Equation to estimate the mean of the population.

Hat matrix In multiple regression, the matrix $H = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}$. This a projection matrix that maps the vector of observed response values into a vector of fitted values by $\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'y} = \mathbf{Hy}$.

Hidden extrapolation An extrapolation is a prediction in a regression analysis that is made at point $(x_1, x_2, \ldots, x_k)$ that is remote from the data used to generate the model. Hidden extrapolation occurs when it is not obvious that the point is remote. This can occur when multicollinearity is present in the data used to construct the model.

Histogram A univariate data display that uses rectangles proportional in area to class frequencies to visually exhibit features of data such as location, variability, and shape.

Homogeneity test In a two-way (r by c) contingency table, this tests if the proportions in the c categories are the same for all r populations.

Homogenous Proportions in c categories are the same for all r populations.

Hypergeometric distribution Distribution where a set of N objects contains K objects classified as successes, $N - K$ objects classfied as failures, and a sample of size n objects is selected randomly (without replacement) from the N objects where $K \leq N$ and $n \leq N$.

Hypergeometric random variable A discrete random variable that is the number of success obtained from a sample drawn without replacement from a finite populations.

Hypothesis (as in statistical hypothesis) A statement about the parameters of a probability distribution or a model, or a statement about the form of a probability distribution.

Hypothesis testing Any procedure used to test a statistical hypothesis.

Identity element A special type of element of a set that leaves other elements unchanged when combined with them.

Independence A property of a probability model and two (or more) events that allows the probability of the intersection to be calculated as the product of the probabilities.

Independence test In a two-way (r by c) contingency table, this tests if the row and column categories are independent.

Indicator variable(s) Variables that are assigned numerical values to identify the levels of a qualitative or categorical response. For example, a response with two categorical levels (yes and no) could be represented with an indicator variable taking on the values 0 and 1.

Individuals control chart A Shewhart control chart in which each plotted point is an individual measurement, rather than a summary statistic. *See* Control chart, Shewhart control chart.

Inference Conclusion from a statistical analysis. It usually refers to the conclusion from a hypothesis test or an interval estimate.

Influential observation An observation in a regression analysis that has a large effect on estimated parameters in the model. Influence is measured by the change in parameters when the influential observation is included and excluded in the analysis.

Interaction In factorial experiments, two factors are said to interact if the effect of one variable is different at different levels of the other variables. In general, when variables operate independently of each other, they do not exhibit interaction.

Interquartile range The difference between the third and first quartiles in a sample of data. The interquartile range is less sensitive to extreme data values than the usual sample range.

Intrinsically linear model In regression analysis, a nonlinear function that can be expressed as a linear function after a suitable transformation is called intrinsically linear.

Joint probability density function A function used to calculate probabilities for two or more continuous random variables.

Joint probability distribution The probability distribution for two or more random variables in a random experiment. *See* Joint probability mass function and Joint probability density function.

Joint probability mass function A function used to calculate probabilities for two or more discrete random variables.

Lack of memory property A property of a Poisson process. The probability of a count in an interval depends only on the length of the interval (and not on the starting point of the interval). A similar property holds for a series of Bernoulli trials. The probability of a success in a specified number of trials depends only on the number of trials (and not on the starting trial).

Large sample CI Confidence interval on the difference in mean where the population variances are unknown.

Large-sample test Sample variances are substituted into the test statistic where the population variances are unknown.

Least squares (method of) A method of parameter estimation in which the parameters of a system are estimated by minimizing the sum of the squares of the differences between the observed values and the fitted or predicted values from the system.

Least squares estimator Any estimator obtained by the method of least squares.

Least squares normal equation Normal equation derived by the method of least squares.

Levels of a factor The settings (or conditions) used for a factor in an experiment.

Leverage, leverage point A point unusual in its location in the x-space that may be influential.

Likelihood function Suppose that the random variables $X_1, X_2, \ldots, X_n$ have a joint distribution given by $f(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots, \theta_p)$ where the θs are unknown parameters. This joint distribution, considered as a function of the θs for fixed x's, is called the likelihood function.

Linear combination An expression constructed from a set of terms by multiplying each term by a constant and adding the results.

Linear function of random variables A random variable that is defined as a linear function of several random variables.

Linear relationship Correlation between 2 variables or 2 sets of data.

Log likelihood surface Surface of graph generated by log likelihood.

Logistic regression A regression model that is used to model a categorical response. For a binary (0, 1) response, the model assumes that the logarithm of the ratio of probabilities (for zero and one) is linearly related to the regressor variables.

Lognormal distribution The distibution of X when the expoent is a random variable W, and W has a normal distribution.

Lognormal random variable A continuous random variable with probability distribution equal to that of $\exp(W)$ for a normal random variable W.

Main effect An estimate of the effect of a factor (or variable) that independently expresses the change in response due to a change in that factor, regardless of other factors that may be present in the system.

Marginal probability density function The probability density function of a continuous random variable obtained from the joint probability distribution of two or more random variables.

Marginal probability distribution The probability distribution of a random variable obtained from the joint probability distribution of two or more random variables.

Maximum likelihood estimator A method of parameter estimation that maximizes the likelihood function of a sample.

Mean The mean usually refers either to the expected value of a random variable or to the arithmetic average of a set of data.

Mean square In general, a mean square is determined by dividing a sum of squares by the number of degrees of freedom associated with the sum of squares.

Mean square(d) error The expected squared deviation of an estimator from the true value of the parameter it estimates. The mean square error can be decomposed into the variance of the estimator plus the square of the bias; that is, $\mathbf{MSE}(\hat{\boldsymbol{\Theta}}) = \mathbf{E}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\theta})^2 = \mathbf{V}(\hat{\boldsymbol{\Theta}}) + [\mathbf{E}(\hat{\boldsymbol{\Theta}}) - \boldsymbol{\theta}]^2$.

Mean—continuous random variable $\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$

Mean—function of a continuous random variable $E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx$

Mechanistic model A model developed from theoretical knowledge or experience in contrast to a model developed from data. *See* Empirical model.

Median The median of a set of data is that value that divides the data into two equal halves. When the number of observations is even, say $2n$, it is customary to define the median as the average of the nth and $(n + 1)$st rank-ordered values. The median can also be defined for a random variable. For example, in the case of a continuous random variable X, the median M can be defined as $\int_{-\infty}^{M} f(x)dx = \int_{M}^{\infty} f(x)dx = 1/2$.

Meta-analysis Combining results from several studies or experimens for analysis.

Minimax criterion Criterion selectes the action from the minimum and maximum, focusing on the worst-case scenario.

Minimin criterion Criterion selectes the action from the minimum, focusing on the best-case scenario.

Minimum variance unbiased estimator (MVUE) The unbiased estimator with the smalled variance.

Model adequacy Suitability of a model for use in an experiment.

Model matrix The X matrix; an $(n \times p)$ matrix of the levels of the independent variables.

Moment about the origin The expected value of a function of a random variable such as $E(X - c)r$ for constants c and r, when $c = 0$.

Moment estimator A method of estimating parameters by equating sample moments to population moments. Since the population moments will be functions of the unknown parameters, this results in equations that may be solved for estimates of the parameters.

Moment-generating function A function that is used to determine properties (such as moments) of the probability distribution of a random variable. It is the expected value of $\exp(tX)$. *See* Generating function and Moment.

Most probable criterion Criterion selects the action to minimize the cost of the most probable state.

Moving range The absolute value of the difference between successive observations in time-ordered data. Used to estimate chance variation in an individual control chart.

Multicollinearity A condition occurring in multiple regression where some of the predictor or regressor variables are nearly linearly dependent. This condition can lead to instability in the estimates of the regression model parameters.

Multinomial distribution The joint probability distribution of the random variables that count the number of results in each of k classes in a random experiment with a series of independent trials with constant probability of each class on each trial. It generalizes a binomial distribution.

Multiple comparisons methods Methods for investigating which means are different.

Multiple regression model A regression model that contains more than one regressor variable.

Multiplication rule For probability, a formula used to determine the probability of the intersection of two (or more) events. For counting techniques, a formula used to determine the number of ways to complete an operation from the number of ways to complete successive steps.

Mutually exclusive events A collection of events whose intersections are empty.

Natural tolerance limits A set of symmetric limits that are three times the process standard deviation from the process mean.

Negative binomial distribution A generalization of a geometric distribution in which the random variable is the numberof Bernoulli trials required to obtain r successes.

Normal approximation A method to approximate probabilities for binomial and Poisson random variables.

Normal equations The set of simultaneous linear equations arrived at in parameter estimation using the method of least squares.

Normal probability plot A specially constructed plot for a variable x (usually on the abscissa) in which y (usually on the ordinate) is scaled so that the graph of the normal cumulative distribution is a straight line.

Normal random variable A continuous random variable that is the most important one in statistics because it results from the central limit theorem. *See* Central limit theorem.

***NP* chart** An attribute control chart that plots the total of defective units in a subgroup. Similar to a fraction-defective chart or P chart.

Nuisance factor A factor that probably influences the response variable, but which is of no interest in the current study. When the levels of the nuisance factor can be controlled, blocking is the design technique that is customarily used to remove its effect.

Null distribution In a hypothesis test, the distribution of the test statistic when the null hypothesized is assumed to be true.

Null hypothesis This term generally relates to a particular hypothesis that is under test, as distinct from the alternative hypothesis (which defines other conditions that are feasible but not being tested). The null hypothesis determines the probability of type I error for the test procedure.

Observational study A system is observed and data might be collected, but changes are not made to the system. *See* Experiment.

Observed significance level *See P*-value.

Odds ratio The odds equals the ratio of two probabilities. In logistic regression, the logarithm of the odds is modeled as a linear function of the regressors. Given values for the regressors at a point, the odds can be calculated. The odds ratio is the odds at one point divided by the odds at another.

One-sided alternative hypothesis A hypothesis other than the one that is being tested where the outcome is fixed in advance.

One-sided confidence bounds Only the lower or upper confidence limits are used for a confidence interval.

Operating characteristic curves (OC curves) A plot of the probability of type II error versus some measure of the extent to which the null hypothesis is false. Typically, one OC curve is used to represent each sample size of interest.

Optimization experiment A experiment conducted to improve (or optimize) a system or process. It is assumed that the important factors are known.

Orthogonal There are several related meanings, including the mathematical sense of perpendicular, two variables being said to be orthogonal if they are statistically independent, or in experimental design where a design is orthogonal if it admits statistically independent estimates of effects.

Orthogonal design *See* Orthogonal.

Outcome An element of a sample space.

Outlier(s) One or more observations in a sample that are so far from the main body of data that they give rise to the question that they may be from another population.

Overall mean Parameter common to all treatments.

Overcontrol Unnecessary adjustments made to processes that increase the deviations from target.

***P* chart** An attribute control chart that plots the proportion of defective units in a subgroup. Also called a fraction-defective control chart. Similar to an *NP* chart.

***P*-value** The exact significance level of a statistical test; that is, the probability of obtaining a value of the test statistic that is at least as extreme as that observed when the null hypothesis is true.

Paired *t*-test Data collected in pairs is analyzed during a test procedure to test the differences.

Parameter estimation The process of estimating the parameters of a population or probability distribution. Parameter estimation, along with hypothesis testing, is one of the two major techniques of statistical inference.

Parametric methods Based on a particular parametric family of distributions.

Pareto chart A bar chart used to rank the causes of a problem.

Pareto diagram A bar chart used to rank the causes of a problem.

Partial regression coefficients Parameters that measure expected change in a regression model.

Partial test; also marginal test Regression coefficent depends on all other regressor variables in a model.

Pattern recognition Recognizing systematic or nonrandom patterns on control charts.

Percentile The set of values that divide the sample into 100 equal parts.

Permutation An ordered sequence of the elements in a set used to determine the number of outcomes in events and sample spaces.

Point estimate A reasonable value of a parameter; a single numerical value θ of a statistic θ.

Point estimator *See* Estimator.

Poisson distribution The random variable X that equals the number of events in a Poisson process is a Poisson random variable with parameter $0 < \lambda$.

Poisson process A random experiment with events that occur in an interval and satisfy the following assumptions. The interval can be partitioned into subintervals such that the probability of more than one event in a subinterval is zero, the probability of an event in a subinterval is proportional to the length of the subinterval, and the event in each subinterval is independent of other subintervals.

Poisson random variable A discrete random variable that is the number of events that occur in a Poisson process.

Polynomial regression model Linear regression models that are equations using polynomials.

Pooled estimator Combining samples generated from the same model to estimate a variance.

Pooled *t*-test A hypothesis to compare the means of two populations with the variances assumed to be equal.

Population Any finite or infinite collection of individual units or objects.

Population or distribution moments The expected value of a function of a random variable such as $E(X - c)^r$ for constants c and r. When $c = 0$, it is said that the moment is about the origin. *See* Moment generating function.

Population standard deviation *See* Standard deviation.

Population variance *See* Variance.

Posterior distribution The probability distribution for a parameter in a Bayesian analysis calculated from the prior distribution and the conditional distribution of the data given the parameter.

Power The power of a statistical test is the probability that the test rejects the null hypothesis when the null hypothesis is indeed false. Thus, the power is equal to one minus the probability of type II error.

Power function A function that describes the relationships among the power of a statistical test, the sample size, and the value of the parameter of interest.

Practical significance Little to no significance on real events; also engineering significance.

Predicted residual sum of squares *See* PRESS statistic.

Prediction interval The interval between a set of upper and lower limits associated with a predicted value designed to show on a probability basis the range of error associated with the prediction.

PRESS statistic In regression analysis, the predicted residual sum of squares. Delete each point and estimate the parameters of the model from the data that remain. Estimate the deleted point from this model. Restore the point and then delete the next point. Each point is estimated once and the sum of squares of these errors is calculated.

Principal block The block containing the treatment combination.

Principal fraction The fraction with the plus sign in a defining relation.

Prior distribution The initial probability distribution assumed for a parameter in a Bayesian analysis.

Probabiiity distribution—continuous random variable A function that provides probabilities in the range of a continuous random variable.

Probability A numerical measure between 0 and 1 assigned to events in a sample space. Higher numbers indicate the event is more likely to occur. *See* Axioms of probability.

Probability density function A function used to calculate probabilities and to specify the probability distribution of a continuous random variable.

Probability distribution For a sample space, a description of the set of possible outcomes along with a method to determine probabilities. For a random variable, a probability distribution is a description of the range along with a method to determine probabilities.

Probability mass function A function that provides probabilities for the values in the range of a discrete random variable.

Process capability The capability of a process to produce product within specification limits. *See* Process capability ratio, Process capability study, *PCR*, and PCR_k.

Process capability ratio A ratio that relates the width of the product specification limits

to measures of process performance. Used to quantify the capability of the process to produce product within specifications. *See* Process capability, Process capability study, *PCR*, and PCR_k.

Projection property A design that projects into a full factorial in any two of the three original factors.

Quality improvement One aspect of quality management, a set of guidelines to ensure that an organization, product, or service is consistent.

Quartiles The three values of a variable that partition it into four equal parts. The central value is usually called the median and the lower and upper values are usually called the lower and upper quartiles, respectively. *See* Quantiles.

***R* chart** A control chart that plots the range of the measurements in a subgroup that is used to monitor the variance of the process.

R^2 A quantity used in regression models to measure the proportion of total variability in the response accounted for by the model. Computationally, $R^2 = SS_{\text{Regression}}/SS_{\text{Total}}$, and large values of R^2 (near unity) are considered good. However, it is possible to have large values of R^2 and find that the model is unsatisfactory. R^2 is also called the coefficient of determination (or the coefficient of multiple determination in multiple regression).

Raleigh distribution A special case of the Weibull distribution when the shape parameter is 2.

Random experiment An experiment that can result in different outcomes, even though it is repeated in the same manner each time.

Random factor In analysis of variance, a factor whose levels are chosen at random from some population of factor levels.

Random interval End-points involve a random variable.

Random sample A sample is said to be random if it is selected in such a way so that every possible sample has the same probability of being selected.

Random variable A function that assigns a real number to each outcome in the sample space of a random experiment.

Randomization Randomly assign treatments to experimental units or conditions in an experiment. This is done to reduce the opportunity for a treatment to be favored or disfavored (biased) by test conditions.

Randomized complete block design A type of experimental design in which treatment or factor levels are assigned to blocks in a random manner.

Rank In the context of data, the rank of a single observation is its ordinal number when all data values are ordered according to some criterion, such as their magnitude.

Rational subgroup A sample of data selected in a manner to include chance sources of variation and to exclude assignable sources of variation to the extent possible.

Reference distribution The distribution of a test statistic when the null hypothesis is true. Sometimes a reference distribution is called the null distribution of the test statistic.

Regression coefficient(s) The parameter(s) in a regression model.

Regression line (or curve) A graphical display of a regression model, usually with the response y on the ordinate and the regressor x on the abscissa.

Regression sum of squares The portion of the total sum of squares attributable to the model that has been fit to the data.

Rejection region In hypothesis testing, this is the region in the sample space of the test statistic that leads to rejection of the null hypothesis when the test statistic falls in this region.

Relative efficiency Comparison of 2 estimators derived by dividing the mean squared errors by each other.

Relative frequency distribution The relative frequency of an event is the proportion of times the event occurred in a series of trials of a random experiment.

Replicates One of the independent repetitions of one or more treatment combinations in an experiment.

Reproductive property of the normal distribution A linear combination of independent, normal random variables is a normal random variable.

Residual Generally this is the difference between the observed and the predicted value of some variable. For example, in regression a residual is the difference between the observed value of the response and the corresponding predicted value obtained from the regression model.

Residual analysis (and plots) Any technique that uses the residuals, usually to investigate the adequacy of the model that was used to generate the residuals.

Resolution A measure of severity of aliasing in a fractional factorial design. We commonly consider resolution III, IV, and V designs.

Response (variable) The dependent variable in a regression model or the observed output variable in a designed experiment.

Response surface When a response y depends on a function of k quantitative variables $x_1, x_2, \ldots, x_k$, the values of the response may be viewed as a surface in $k + 1$ dimensions. This surface is called a response surface. Response surface methodology is a subset of experimental design concerned with approximating this surface with a model and using the resulting model to optimize the system or process.

Ridge regression Alternative to ordinary least squares useful in combatting multicollinearity.

Run Arrangement of points in a distinct pattern.

Run rules A set of rules applied to the points plotted on a Shewhart control chart that are used to make the chart more sensitized to assignable causes. *See* Control chart, Shewhart control chart.

***S* chart** A control chart that plots the standard deviation of the measurements in a subgroup that is used to monitor the variance of the process.

Sample Any subset of the elements of a population.

Sample Mean The arithmetic average or mean of the observations in a sample. If the observations are $x_1, x_2, \ldots, x_n$, then the sample mean is $(1/n)g_{i=1}^{n}x_i$. The sample mean is usually denoted by $\bar{x}$.

Sample median The median of a set of data is that value that divides the data into two equal halves. When the number of observations is even, say $2n$, it is customary to define the median as the average of the nth and $(n + 1)$st rank-ordered values. The median can also be defined for a random variable. For example, in the case of a continuous random variable X, the median M can be defined as $\int_{-\infty}^{M} f(x)dx = \int_{M}^{\infty} f(x)dx = 1/2$.

Sample moment The quantity $(1/n)\, g_{i=1}^{n}\, x_i^k$ is called the kth sample moment.

Sample range *See* Range.

Sample space The set of all possible outcomes of a random experiment.

Sample standard deviation The positive square root of the sample variance. The sample standard deviation is the most widely used measure of variability of sample data.

Sample variance A measure of variability of sample data, defined as $s^2 = [1/(n-1)]\mathbf{g}_{i=1}^{n}(x_i - \bar{x})^2$, where $\bar{x}$ is the sample mean.

Sampling distribution The probability distribution of a statistic. For example, the sampling distribution of the sample mean $\overline{\mathbf{X}}$ is the normal distribution.

Saturated fractional factorial Design where all available degrees of freedom are used to estimate main effects.

Scale parameter A parameter that defines the dispersion of a sample or a probability distribution.

Scatter diagram A diagram displaying observations on two variables, x and y. Each observation is represented by a point showing its x-y coordinates. The scatter diagram can be very effective in revealing the joint variability of x and y or the nature of the relationship between them.

Screening experiment An experiment designed and conducted for the purpose of screening out or isolating a promising set of factors for future experimentation. Many screening experiments are fractional factorials, such as two-level fractional factorial designs.

Shewhart control chart A specific type of control chart developed by Walter A. Shewhart. Typically, each plotted point is a summary statistic calculated from the data in a rational subgroup. *See* Control chart.

Sign test A statistical test based on the signs of certain functions of the observations and not their magnitudes.

Significance level *See* Level of significance.

Six-sigma process Originally used to describe a process with the mean at least six standard deviations from the nearest specification limits. It has now been used to describe any process with a defect rate of 3.4 parts per million.

Sparsity of effects principle The system is usually dominated by the main effects and low-order interactions.

Specification limits Numbers that define the region of measurement for acceptable product. Usually there is an upper and lower limit, but one-sided limits can also be used.

Standard deviation The positive square root of the variance. The standard deviation is the most widely used measure of variability.

Standard deviation—continuous random variable $\sigma = \sqrt{\sigma^2}$.

Standard error The standard deviation of the estimator of a parameter. The standard error is also the standard deviation of the sampling distribution of the estimator of a parameter.

Standard normal random variable A normal random variable with mean zero and variance one that has its cumulative distribution function tabulated in Appendix Table II.

Standardize The transformation of a normal random variable that subtracts its mean and divides by its standard deviation to generate a standard normal random variable.

Standardized residual In regression, the standardized residual is computed by dividing the ordinary residual by the square root of the residual mean square. This produces scaled residuals that have, approximately, a unit variance.

Statistic A summary value calculated from a sample of observations. Usually, a statistic is an estimator of some population parameter.

Statistical hypothesis A statement about the parameters of one or more populations.

Statistical inference *See* Inference.

Statistical Process Control (SPC) A set of problem-solving tools based on data that are used to improve a process.

Statistical quality control Statistical and engineering methods used to measure, monitor, control, and improve quality.

Statistical significance *See* Significance.

Statistics The science of collecting, analyzing, interpreting, and drawing conclusions from data.

Steepest ascent (or descent) A strategy for a series of tests to optimize a response used along with response surface models.

Stem-and-leaf diagram A method of displaying data in which the stem corresponds to a range of data values and the leaf represents the next digit. It is an alternative to the histogram but displays the individual observations rather than sorting them into bins.

Stepwise regression and related methods A method of selecting variables for inclusion in a regression model. It operates by introducing the candidate variables one at a time (as in forward selection) and then attempting to remove variables following each forward step.

Studentized residual In regression, the studentized residual is calculated by dividing the ordinary residual by its exact standard deviation, producing a set of scaled residuals that have, exactly, unit standard deviation.

Sum of squares identity Partition of the total sum of squares.

***t* distribution** The distribution of the random variable defined as the ratio of two independent random variables. The numerator is a standard normal random variable and the denominator is the square root of a chi-square random variable divided by its number of degrees of freedom.

***t*-test** Any test of significance based on the t distribution. The most common t-tests are (1) testing hypotheses about the mean of a normal distribution with unknown variance, (2) testing hypotheses about the means of two normal distributions, and (3) testing hypotheses about individual regression coefficients.

Tampering Another name for overcontrol.

Test Statistic A function of a sample of observations that provides the basis for testing a statistical hypothesis.

Three-dimensional surface plot Displays showing results of experiments in three dimensions.

Time series A set of ordered observations taken at points in time.

Tolerance chart A grapical tool displaying all tolerance data.

Tolerance interval An interval that contains a specified proportion of a population with a stated level of confidence.

Total probability rule Given a collection of mutually exclusive events whose union is the sample space, the probability of an event can be written as the sum of the probabilities of the intersections of the event with the members of this collection.

Total sum of squares Equation that describes total variability in data.

Treatment In experimental design, a treatment is a specific level of a factor of interest. Thus, if the factor is temperature, the treatments are the specific temperature levels used in the experiment.

Treatment sum of squares In analysis of variance, this is the sum of squares that accounts for the variability in the response variable due to the different treatments that have been applied.

Two one-sided tests (TOST) Apply two tests to the same sample of data.

Two-factor interaction plots Displays showing interactions in results of experiments with two factors.

Two-sided alternative hypothesis A hypothesis other than the one that is being tested, where there is some effect.

Type I Error In hypothesis testing, an error incurred by rejecting a null hypothesis when it is actually true (also called an $\boldsymbol{\alpha}$-error).

Type II Error In hypothesis testing, an error incurred by failing to reject a null hypothesis when it is actually false (also called a $\boldsymbol{\beta}$-error).

***U* chart** An attribute control chart that plots the average number of defects per unit in a subgroup. Also called a defects-per-unit control chart. Similar to a C chart.

Unbiased estimator An estimator that has its expected value equal to the parameter that is being estimated is said to be unbiased.

Unbiased point estimator *See* Unbiased estimator.

Unreplicated design A single replicate of a statistical design.

Variable selection The problem of selecting a subset of variables for a model from a candidate list that contains all or most of the useful information about the response in the data.

Variables control chart Any control chart for a continuous random variable. *See* Attribute control chart.

Variance A measure of variability defined as the expected value of the square of the random variable around its mean.

Variance—continuous random variable

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

$$= \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

Variance component In analysis of variance models involving random effects, one of the objectives is to determine how much variability can be associated with each of the potential sources of variability defined by the experimenters. It is customary to define a variance associated with each of these sources. These variances in some sense sum to the total variance of the response, and are usually called variance components.

Variance inflation factors Quantities used in multiple regression to assess the extent of multicollinearity (or near linear dependence) in the regressors. The variance inflation factor for the ith regressor VIF_i can be defined as $VIF_i = [1/(1 - R_i^2)]$, where R_i^2 is the coefficient of determination obtained when x_i is regressed on the other regressor variables. Thus, when x_i is nearly linearly dependent on a subset of the other regressors, R_i^2 will be close to unity and the value of the corresponding variance inflation factor will be large. Values of the variance inflation factors that exceed 10 are usually taken as a signal that multicollinearity is present.

Warning limits Horizontal lines added to a control chart (in addition to the control limits) that are used to make the chart more sensitive to assignable causes.

Weibull random variable A continuous random variable that is often used to model the time until failure of a physical system. The parameters of the distribution are flexible enough that the probability density function can assume many different shapes.

Weighted average Average of two sample variances where the weights depend on two sample sizes.

Western Electric rules A specific set of run rules that were developed at Western Electric Corporation. *See* Run rules.

Wilcoxon rank-sum test A nonparametric test for the equality of means in two populations. This is sometimes called the Mann-Whitney test.

Wilcoxon signed-rank test A distribution-free test of the equality of the location parameters of two otherwise identical distributions. It is an alternative to the two-sample t-test for nonnormal populations.

With replacement A method to select samples in which items are replaced between successive selections.

Without replacement A method to select samples in which items are *not* replaced between successive selections.

$\bar{X}$ chart A control chart that plots the average of the measurements in a subgroup that is used to monitor the process mean.

z-test Test for the standard normal distribution.

z-value Value associated with a probability, obtained by standardizing X.

Exercises

Chapter 2 Exercises

Exercises for Section 2.1

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

Provide a reasonable description of the sample space for each of the random experiments in Exercises 2.1.1 to 2.1.11. There can be more than one acceptable interpretation of each experiment. Describe any assumptions you make.

2.1.1 Each of four transmitted bits is classified as either in error or not in error.

2.1.2 The number of hits (views) is recorded at a high-volume Web site in a day.

2.1.3 In the final inspection of electronic power supplies, either units pass or three types of nonconformities might occur: functional, minor, or cosmetic. Three units are inspected.

2.1.4 An ammeter that displays three digits is used to measure current in milliamperes.

2.1.5 WP The following two questions appear on an employee survey questionnaire. Each answer is chosen from the five-point scale 1 (never), 2, 3, 4, 5 (always).

Is the corporation willing to listen to and fairly evaluate new ideas?

How often are my coworkers important in my overall job performance?

2.1.6 The time until a service transaction is requested of a computer to the nearest millisecond.

2.1.7 SS The pH reading of a water sample to the nearest tenth of a unit.

2.1.8 The voids in a ferrite slab are classified as small, medium, or large. The number of voids in each category is measured by an optical inspection of a sample.

2.1.9 A sampled injection-molded part could have been produced in either one of two presses and in any one of the eight cavities in each press.

2.1.10 An order for an automobile can specify either an automatic or a standard transmission, either with or without air conditioning, and with any one of the four colors red, blue, black, or white. Describe the set of possible orders for this experiment.

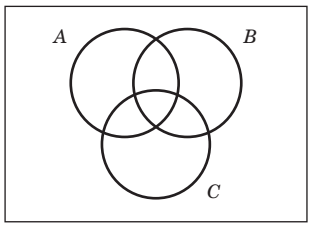
2.1.11 WP Calls are repeatedly placed to a busy phone line until a connection is achieved.

2.1.12 Three attempts are made to read data in a magnetic storage device before an error recovery procedure is used. The error recovery procedure attempts three corrections before an "abort" message is sent to the operator. Let

s denote the success of a read operation
f denote the failure of a read operation
S denote the success of an error recovery procedure
F denote the failure of an error recovery procedure
A denote an abort message sent to the operator

Describe the sample space of this experiment with a tree diagram.

2.1.13 SS VS Three events are shown on the Venn diagram in the following figure:



Reproduce the figure and shade the region that corresponds to each of the following events.

a. A' **b.** $A \cap B$ **c.** $(A \cap B) \cup C$
d. $(B \cup C)'$ **e.** $(A \cap B)' \cup C$

2.1.14 In an injection-molding operation, the length and width, denoted as X and Y, respectively, of each molded part are evaluated. Let

A denote the event of $48 < X < 52$ centimeters
B denote the event of $9 < Y < 11$ centimeters

Construct a Venn diagram that includes these events. Shade the areas that represent the following:

a. A **b.** $A \cap B$
c. $A' \cup B$ **d.** $A \cap B$

e. If these events were mutually exclusive, how successful would this production operation be? Would the process produce parts with $X = 50$ centimeters and $Y = 10$ centimeters?

2.1.15 WP A digital scale that provides weights to the nearest gram is used.

a. What is the sample space for this experiment?

Let A denote the event that a weight exceeds 11 grams, let B denote the event that a weight is less than or equal to 15 grams,

and let C denote the event that a weight is greater than or equal to 8 grams and less than 12 grams.
Describe the following events.

b. $A \cup B$ **c.** $A \cap B$
d. A' **e.** $A \cup B \cup C$
f. $(A \cup C)'$ **g.** $A \cap B \cap C$
h. $B' \cap C$ **i.** $A \cup (B \cap C)$

2.1.16 In light-dependent photosynthesis, light quality refers to the wavelengths of light that are important. The wavelength of a sample of photosynthetically active radiations (PAR) is measured to the nearest nanometer. The red range is 675–700 nm and the blue range is 450–500 nm. Let A denote the event that PAR occurs in the red range, and let B denote the event that PAR occurs in the blue range. Describe the sample space and indicate each of the following events:

a. A **b.** B **c.** $A \cap B$ **d.** $A \cup B$

2.1.17 WP GO Tutorial Four bits are transmitted over a digital communications channel. Each bit is either distorted or received without distortion. Let A_i denote the event that the ith bit is distorted, $i = 1, \ldots, 4$.

a. Describe the sample space for this experiment.
b. Are the A_i's mutually exclusive?

Describe the outcomes in each of the following events:

c. A_1 **d.** A_1'
e. $A_1 \cap A_2 \cap A_3 \cap A_4$ **f.** $(A_1 \cap A_2) \cup (A_3 \cap A_4)$

2.1.18 WP Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized here:

		Shock Resistance	
		High	Low
Scratch Resistance	High	70	9
	Low	16	5

Let A denote the event that a disk has high shock resistance, and let B denote the event that a disk has high scratch resistance. Determine the number of disks in $A \cap B$, A', and $A \cup B$.

2.1.19 SS In control replication, cells are replicated over a period of two days. Two control mechanisms have been identified—one positive and one negative. Suppose that a replication is observed in three cells. Let A denote the event that all cells are identified as positive, and let B denote the event that all cells are negative. Describe the sample space graphically and display each of the following events:

a. A **b.** B **c.** $A \cap B$ **d.** $A \cup B$

2.1.20 WP Samples of emissions from three suppliers are classified for conformance to air-quality specifications. The results from 100 samples are summarized as follows:

		Conforms	
		Yes	No
Supplier	1	22	8
	2	25	5
	3	30	10

Let A denote the event that a sample is from supplier 1, and let B denote the event that a sample conforms to specifications. Determine the number of samples in $A' \cap B$, B', and $A \cup B$.

2.1.21 WP The rise time of a reactor is measured in minutes (and fractions of minutes). Let the sample space be positive, real numbers. Define the events A and B as follows: $A = \{x \mid x < 72.5\}$ and $B = \{x \mid x > 52.5\}$.
Describe each of the following events:

a. A' **b.** B'
c. $A \cap B$ **d.** $A \cup B$

2.1.22 The following table summarizes 204 endothermic reactions involving sodium bicarbonate.

Final Temperature Conditions	Heat Absorbed (cal)	
	Below Target	Above Target
266 K	12	40
271 K	44	16
274 K	56	36

Let A denote the event that a reaction's final temperature is 271 K or less. Let B denote the event that the heat absorbed is below target. Determine the number of reactions in each of the following events.

a. $A \cap B$ **b.** A' **c.** $A \cup B$
d. $A \cup B'$ **e.** $A' \cap B'$

2.1.23 SS A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. How many different designs are possible?

2.1.24 Consider the hospital emergency department data in Example 2.6. Let A denote the event that a visit is to hospital 1, and let B denote the event that a visit results in admittance to any hospital. Determine the number of persons in each of the following events.

a. $A \cap B$ **b.** A' **c.** $A \cup B$
d. $A \cup B'$ **e.** $A' \cap B'$

2.1.25 The article "Term Efficacy of Ribavirin Plus Interferon Alfa in the Treatment of Chronic Hepatitis C" [*Gastroenterology* (1996, Vol. 111(5), pp. 1307–1312)], considered the effect of two treatments and a control for treatment of hepatitis C. The following table provides the total patients in each group and the number that showed a complete (positive) response after 24 weeks of treatment.

	Complete Response	Total
Ribavirin plus interferon alfa	16	21
Interferon alfa	6	19
Untreated controls	0	20

Let A denote the event that the patient was treated with ribavirin plus interferon alfa, and let B denote the event that the response was complete. Determine the number of patients in each of the following events.

a. A **b.** $A \cap B$ **c.** $A \cup B$ **d.** $A' \cap B'$

2.1.26 A computer system uses passwords that contain exactly eight characters, and each character is 1 of the 26 lowercase letters (a–z) or 26 uppercase letters (A–Z) or 10 integers (0–9). Let Ω denote the set of all possible passwords, and let A and B denote the events that consist of passwords with only letters or only integers, respectively. Determine the number of passwords in each of the following events.

a. Ω **b.** A **c.** $A' \cap B'$
d. Passwords that contain at least 1 integer
e. Passwords that contain exactly 1 integer

Exercises for Section 2.2

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

2.2.1 WP A sample of two printed circuit boards is selected without replacement from a batch. Describe the (ordered) sample space for each of the following batches:

a. The batch contains 90 boards that are not defective, 8 boards with minor defects, and 2 boards with major defects.
b. The batch contains 90 boards that are not defective, 8 boards with minor defects, and 1 board with major defects.

2.2.2 A sample of two items is selected without replacement from a batch. Describe the (ordered) sample space for each of the following batches:

a. The batch contains the items $\{a, b, c, d\}$.
b. The batch contains the items $\{a, b, c, d, e, f, g\}$.
c. The batch contains 4 defective items and 20 good items.
d. The batch contains 1 defective item and 20 good items.

2.2.3 WP SS A wireless garage door opener has a code determined by the up or down setting of 12 switches. How many outcomes are in the sample space of possible codes?

2.2.4 WP In a manufacturing operation, a part is produced by machining, polishing, and painting. If there are three machine tools, four polishing tools, and three painting tools, how many different routings (consisting of machining, followed by polishing, and followed by painting) for a part are possible?

2.2.5 WP New designs for a wastewater treatment tank have proposed three possible shapes, four possible sizes, three locations for input valves, and four locations for output valves. How many different product designs are possible?

2.2.6 WP A manufacturing process consists of 10 operations that can be completed in any order. How many different production sequences are possible?

2.2.7 WP VS A batch of 140 semiconductor chips is inspected by choosing a sample of 5 chips without replacement. Assume 10 of the chips do not conform to customer requirements.

a. How many different samples are possible?
b. How many samples contain exactly one nonconforming chip?
c. How many samples contain at least one nonconforming chip?

2.2.8 WP In a sheet metal operation, three notches and four bends are required. If the operations can be done in any order, how many different ways of completing the manufacturing are possible?

2.2.9 In the laboratory analysis of samples from a chemical process, five samples from the process are analyzed daily. In addition, a control sample is analyzed twice each day to check the calibration of the laboratory instruments.

a. How many different sequences of process and control samples are possible each day? Assume that the five process samples are considered identical and that the two control samples are considered identical.
b. How many different sequences of process and control samples are possible if we consider the five process samples to be different and the two control samples to be identical?
c. For the same situation as part (b), how many sequences are possible if the first test of each day must be a control sample?

2.2.10 In the layout of a printed circuit board for an electronic product, 12 different locations can accommodate chips.

a. If five different types of chips are to be placed on the board, how many different layouts are possible?
b. If the five chips that are placed on the board are of the same type, how many different layouts are possible?

2.2.11 SS Consider the design of a communication system.

a. How many three-digit phone prefixes that are used to represent a particular geographic area (such as an area code) can be created from the digits 0 through 9?
b. As in part (a), how many three-digit phone prefixes are possible that do not start with 0 or 1, but contain 0 or 1 as the middle digit?
c. How many three-digit phone prefixes are possible in which no digit appears more than once in each prefix?

2.2.12 In the design of an electromechanical product, 12 components are to be stacked into a cylindrical casing in a manner that minimizes the impact of shocks. One end of the casing is designated as the bottom and the other end is the top.

a. If all components are different, how many different designs are possible?
b. If 7 components are identical to one another, but the others are different, how many different designs are possible?
c. If 3 components are of one type and identical to one another, and 4 components are of another type and identical to one another, but the others are different, how many different designs are possible?

2.2.13 A bin of 50 parts contains 5 that are defective. A sample of 10 parts is selected without replacement. How many samples contain at least 4 defective parts?

2.2.14 WP Plastic parts produced by an injection-molding operation are checked for conformance to specifications. Each tool contains 12 cavities in which parts are produced, and these parts

fall into a conveyor when the press opens. An inspector selects 3 parts without replacement from among the 12 at random. Two cavities are affected by a temperature malfunction that results in parts that do not conform to specifications.

a. How many samples contain exactly one nonconforming part?

b. How many samples contain at least one nonconforming part?

2.2.15 Similar to the hospital schedule in Example 2.9, suppose that an operating room needs to handle three knee, four hip, and five shoulder surgeries.

a. How many different sequences are possible?

b. How many different sequences have all hip, knee, and shoulder surgeries scheduled consecutively?

c. How many different schedules begin and end with a knee surgery?

Exercises for Section 2.3

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

2.3.1 **WP** The sample space of a random experiment is $\{a, b, c, d, e\}$ with probabilities 0.1, 0.1, 0.2, 0.4, and 0.2, respectively. Let A denote the event $\{a, b, c\}$, and let B denote the event $\{c, d, e\}$. Determine the following:

a. $P(A)$ **b.** $P(B)$ **c.** $P(A')$
d. $P(A \cup B)$ **e.** $P(A \cap B)$

2.3.2 A part selected for testing is equally likely to have been produced on any one of six cutting tools.

a. What is the sample space?

b. What is the probability that the part is from tool 1?

c. What is the probability that the part is from tool 3 or tool 5?

d. What is the probability that the part is not from tool 4?

2.3.3 **WP** **SS** An injection-molded part is equally likely to be obtained from any one of the eight cavities on a mold.

a. What is the sample space?

b. What is the probability that a part is from cavity 1 or 2?

c. What is the probability that a part is from neither cavity 3 nor 4?

2.3.4 **WP** A credit card contains 16 digits between 0 and 9. However, only 100 million numbers are valid. If a number is entered randomly, what is the probability that it is a valid number?

2.3.5 **WP** In a NiCd battery, a fully charged cell is composed of nickelic hydroxide. Nickel is an element that has multiple oxidation states and that is usually found in the following states:

Nickel Charge	Proportions Found
0	0.17
+2	0.35
+3	0.33
+4	0.15

a. What is the probability that a cell has at least one of the positive nickel-charged options?

b. What is the probability that a cell is *not* composed of a positive nickel charge greater than +3?

2.3.6 A message can follow different paths through servers on a network. The sender's message can go to one of five servers for the first step; each of them can send to five servers at the second step; each of those can send to four servers at the third step; and then the message goes to the recipient's server.

a. How many paths are possible?

b. If all paths are equally likely, what is the probability that a message passes through the first of four servers at the third step?

2.3.7 **WP** **SS** Suppose your vehicle is licensed in a state that issues license plates that consist of three digits (between 0 and 9) followed by three letters (between A and Z). If a license number is selected randomly, what is the probability that yours is the one selected?

2.3.8 **WP** Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

		Shock Resistance	
		High	Low
Scratch Resistance	High	70	9
	Low	16	5

Let A denote the event that a disk has high shock resistance, and let B denote the event that a disk has high scratch resistance. If a disk is selected at random, determine the following probabilities:

a. $P(A)$ **b.** $P(B)$ **c.** $P(A')$
d. $P(A \cap B)$ **e.** $P(A \cup B)$ **f.** $P(A' \cup B)$

2.3.9 Magnesium alkyls are used as homogenous catalysts in the production of linear low-density polyethylene (LLDPE), which requires a finer magnesium powder to sustain a reaction. Redox reaction experiments using four different amounts of magnesium powder are performed. Each result may or may not be further reduced in a second step using three different magnesium powder amounts. Each of these results may or may not be further reduced in a third step using three different amounts of magnesium powder.

a. How many experiments are possible?

b. If all outcomes are equally likely, what is the probability that the best result is obtained from an experiment that uses all three steps?

c. Does the result in part (b) change if five or six or seven different amounts are used in the first step? Explain.

2.3.10 WP An article in the *Journal of Database Management* ["Experimental Study of a Self-Tuning Algorithm for DBMS Buffer Pools" (2005, Vol. 16, pp. 1–20)] provided the workload used in the TPC-C OLTP (Transaction Processing Performance Council's Version C On-Line Transaction Processing) benchmark, which simulates a typical order entry application.

Average Frequencies and Operations in TPC-C

Transaction	Frequency	Selects	Updates
New order	43	23	11
Payment	44	4.2	3
Order status	4	11.4	0
Delivery	5	130	120
Stock level	4	0	0

The frequency of each type of transaction (in the second column) can be used as the percentage of each type of transaction. The average number of *selects* operations required for each type of transaction is shown. Let A denote the event of transactions with an average number of *selects* operations of 12 or fewer. Let B denote the event of transactions with an average number of *updates* operations of 12 or fewer. Calculate the following probabilities.

a. $P(A)$ **b.** $P(B)$ **c.** $P(A \cap B)$
d. $P(A \cap B')$ **e.** $P(A \cup B)$

2.3.11 VS Samples of emissions from three suppliers are classified for conformance to air-quality specifications. The results from 100 samples are summarized as follows:

		Conforms	
		Yes	No
Supplier	1	22	8
	2	25	5
	3	30	10

Let A denote the event that a sample is from supplier 1, and let B denote the event that a sample conforms to specifications. If a sample is selected at random, determine the following probabilities:

a. $P(A)$ **b.** $P(B)$ **c.** $P(A')$
d. $P(A \cap B)$ **e.** $P(A \cup B)$ **f.** $P(A' \cap B)$

2.3.12 VS Consider the hospital emergency room data in Example 2.6. Let A denote the event that a visit is to hospital 4, and let B denote the event that a visit results in LWBS (at any hospital). Determine the following probabilities.

a. $P(A \cap B)$ **b.** $P(A')$ **c.** $P(A \cup B)$
d. $P(A \cup B')$ **e.** $P(A' \cap B')$

2.3.13 Use the axioms of probability to show the following:

a. For any event E, $P(E') = 1 - P(E)$.
b. $P(\emptyset) = 0$
c. If A is contained in B, then $P(A) \leq P(B)$.

2.3.14 Suppose that a patient is selected randomly from those described in Exercise 2.1.25. Let A denote the event that the patient is in the group treated with interferon alfa, and let B denote the event that the patient has a complete response. Determine the following probabilities.

a. $P(A)$ **b.** $P(B)$ **c.** $P(A \cap B)$
d. $P(A \cup B)$ **e.** $P(A' \cup B)$

2.3.15 SS A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. A specific design is randomly generated by the Web server when you visit the site. If you visit the site five times, what is the probability that you will not see the same design?

2.3.16 Similar to the hospital schedule in Example 2.9, suppose that an operating room needs to schedule three knee, four hip, and five shoulder surgeries. Assume that all schedules are equally likely. Determine the probability for each of the following:

a. All hip surgeries are completed before another type of surgery.
b. The schedule begins with a hip surgery.
c. The first and last surgeries are hip surgeries.
d. The first two surgeries are hip surgeries.

2.3.17 A computer system uses passwords that contain exactly eight characters, and each character is one of 26 lowercase letters (a–z) or 26 uppercase letters (A–Z) or 10 integers (0–9). Let Ω denote the set of all possible passwords, and let A and B denote the events that consist of passwords with only letters or only integers, respectively. Suppose that all passwords in Ω are equally likely. Determine the probability of each of the following:

a. A **b.** B
c. A password contains at least 1 integer.
d. A password contains exactly 2 integers.

Exercises for Section 2.4

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

2.4.1 WP If A, B, and C are mutually exclusive events with $P(A) = 0.2$, $P(B) = 0.3$, and $P(C) = 0.4$, determine the following probabilities:

a. $P(A \cup B \cup C)$ **b.** $P(A \cap B \cap C)$
c. $P(A \cap B)$ **d.** $P[(A \cup B) \cap C]$
e. $P(A' \cap B' \cap C')$

2.4.2 WP If $P(A) = 0.3$, $P(B) = 0.2$, and $P(A \cap B) = 0.1$, determine the following probabilities:

a. $P(A')$ **b.** $P(A \cup B)$ **c.** $P(A' \cap B)$
d. $P(A \cap B')$ **e.** $P[(A \cup B)']$ **f.** $P(A' \cup B)$

2.4.3 WP A manufacturer of front lights for automobiles tests lamps under a high-humidity, high-temperature environment using intensity and useful life as the responses of interest. The following table shows the performance of 130 lamps:

		Useful life	
		Satisfactory	Unsatisfactory
Intensity	Satisfactory	117	3
	Unsatisfactory	8	2

a. Find the probability that a randomly selected lamp will yield unsatisfactory results under any criteria.
b. The customers for these lamps demand 95% satisfactory results. Can the lamp manufacturer meet this demand?

2.4.4 WP In the article "ACL Reconstruction Using Bone-Patellar Tendon-Bone Press-Fit Fixation: 10-Year Clinical Results" in *Knee Surgery, Sports Traumatology, Arthroscopy* (2005, Vol. 13, pp. 248–255), the following causes for knee injuries were considered:

Activity	Percentage of Knee Injuries
Contact sport	46%
Noncontact sport	44%
Activity of daily living	9%
Riding motorcycle	1%

a. What is the probability that a knee injury resulted from a sport (contact or noncontact)?
b. What is the probability that a knee injury resulted from an activity other than a sport?

2.4.5 Consider the hospital emergency room data in Example 2.6. Let A denote the event that a visit is to hospital 4, and let B denote the event that a visit results in LWBS (at any hospital). Use the addition rules to calculate the following probabilities.

a. $P(A \cup B)$ **b.** $P(A \cup B')$ **c.** $P(A' \cup B')$

2.4.6 WP Strands of copper wire from a manufacturer are analyzed for strength and conductivity. The results from 100 strands are as follows:

	Strength	
	High	Low
High conductivity	74	8
Low conductivity	15	3

a. If a strand is randomly selected, what is the probability that its conductivity is high and its strength is high?
b. If a strand is randomly selected, what is the probability that its conductivity is low or its strength is low?
c. Consider the event that a strand has low conductivity and the event that the strand has low strength. Are these two events mutually exclusive?

2.4.7 WP SS A computer system uses passwords that are six characters, and each character is one of the 26 letters (a–z) or 10 integers (0–9). Uppercase letters are not used. Let A denote the event that a password begins with a vowel (either a, e, i, o, or u), and let B denote the event that a password ends with an even number (either 0, 2, 4, 6, or 8). Suppose a hacker selects a password at random. Determine the following probabilities:

a. $P(A)$ **b.** $P(B)$
c. $P(A \cap B)$ **d.** $P(A \cup B)$

2.4.8 A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. A specific design is randomly generated by the Web server when you visit the site. Let A denote the event that the design color is red, and let B denote the event that the font size is not the smallest one. Use the addition rules to calculate the following probabilities.

a. $P(A \cup B)$ **b.** $P(A \cup B')$ **c.** $P(A' \cup B')$

2.4.9 A computer system uses passwords that contain exactly eight characters, and each character is one of the 26 lowercase letters (a–z) or 26 uppercase letters (A–Z) or 10 integers (0–9). Assume all passwords are equally likely. Let A and B denote the events that consist of passwords with only letters or only integers, respectively. Determine the following probabilities:

a. $P(A \cup B)$ **b.** $P(A' \cup B)$
c. P (Password contains exactly 1 or 2 integers)

2.4.10 Consider the three patient groups in Exercise 2.1.25. Let A denote the event that the patient was treated with ribavirin plus interferon alfa, and let B denote the event that the response was complete. Determine the following probabilities:

a. $P(A \cup B)$ **b.** $P(A' \cup B)$ **c.** $P(A \cup B')$

2.4.11 The article "Clinical and Radiographic Outcomes of Four Different Treatment Strategies in Patients with Early Rheumatoid Arthritis" [*Arthritis & Rheumatism* (2005, Vol. 52, pp. 3381–3390)] considered four treatment groups. The groups consisted of patients with different drug therapies (such as prednisone and infliximab): sequential monotherapy (group 1), step-up combination therapy (group 2), initial combination therapy (group 3), or initial combination therapy with infliximab (group 4). Radiographs of hands and feet were used to evaluate disease progression. The number of patients without progression of joint damage was 76 of 114 patients (67%), 82 of 112 patients (73%), 104 of 120 patients (87%), and 113 of 121 patients (93%) in groups 1–4, respectively. Suppose that a patient is selected randomly. Let A denote the event that the patient is in group 1, and let B denote the event that there is no progression. Determine the following probabilities:

a. $P(A \cup B)$ **b.** $P(A' \cup B')$ **c.** $P(A \cup B')$

Exercises for Section 2.5

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

2.5.1 WP SS VS The analysis of results from a leaf transmutation experiment (turning a leaf into a petal) is summarized by type of transformation completed:

		Total Textural Transformation	
		Yes	No
Total color transformation	Yes	243	26
	No	13	18

a. If a leaf completes the color transformation, what is the probability that it will complete the textural transformation?
b. If a leaf does not complete the textural transformation, what is the probability it will complete the color transformation?

2.5.2 WP Samples of skin experiencing desquamation are analyzed for both moisture and melanin content. The results from 100 skin samples are as follows:

		Melanin Content	
		High	Low
Moisture content	High	13	7
	Low	48	32

Let A denote the event that a sample has low melanin content, and let B denote the event that a sample has high moisture content. Determine the following probabilities:

a. $P(A)$ **b.** $P(B)$ **c.** $P(A \mid B)$ **d.** $P(B \mid A)$

2.5.3 WP The following table summarizes the number of deceased beetles under autolysis (the destruction of a cell after its death by the action of its own enzymes) and putrefaction (decomposition of organic matter, especially protein, by microorganisms, resulting in production of foul-smelling matter):

		Autolysis	
		High	Low
Putrefaction	High	14	59
	Low	18	9

a. If the autolysis of a sample is high, what is the probability that the putrefaction is low?
b. If the putrefaction of a sample is high, what is the probability that the autolysis is high?
c. If the putrefaction of a sample is low, what is the probability that the autolysis is low?

2.5.4 WP A maintenance firm has gathered the following information regarding the failure mechanisms for air conditioning systems:

		Evidence of Gas Leaks	
		Yes	No
Evidence of electrical failure	Yes	55	17
	No	32	3

The units without evidence of gas leaks or electrical failure showed other types of failure. If this is a representative sample of AC failure, find the probability

a. That failure involves a gas leak
b. That there is evidence of electrical failure given that there was a gas leak
c. That there is evidence of a gas leak given that there is evidence of electrical failure

2.5.5 Consider the endothermic reactions in Exercise 2.1.22. Let A denote the event that a reaction's final temperature is 271 K or less. Let B denote the event that the heat absorbed is above target. Determine the following probabilities.

a. $P(A \mid B)$ **b.** $P(A' \mid B)$
c. $P(A \mid B')$ **d.** $P(B \mid A)$

2.5.6 WP A batch of 500 containers for frozen orange juice contains 5 that are defective. Three are selected, at random, without replacement from the batch.

a. What is the probability that the second one selected is defective given that the first one was defective?
b. What is the probability that the first two selected are defective?
c. What is the probability that the first two selected are both acceptable?
d. What is the probability that the third one selected is defective given that the first and second ones selected were defective?
e. What is the probability that the third one selected is defective given that the first one selected was defective and the second one selected was okay?
f. What is the probability that all three selected ones are defective?

2.5.7 Suppose A and B are mutually exclusive events. Construct a Venn diagram that contains the three events A, B, and C such that $P(A \mid C) = 1$ and $P(B \mid C) = 0$.

2.5.8 An article in *The Canadian Entomologist* (Harcourt et al., 1977, Vol. 109, pp. 1521–1534) reported on the life of the alfalfa weevil from eggs to adulthood. The following table shows the number of larvae that survived at each stage of development from eggs to adults.

Eggs	Early Larvae	Late Larvae	Pre-pupae	Late Pupae	Adults
421	412	306	45	35	31

a. What is the probability an egg survives to adulthood?
b. What is the probability of survival to adulthood given survival to the late larvae stage?
c. What stage has the lowest probability of survival to the next stage?

2.5.9 SS VS Consider the hospital emergency room data in Example 2.6. Let A denote the event that a visit is to hospital 4, and let B denote the event that a visit results in LWBS (at any hospital). Determine the following probabilities.

a. $P(A \mid B)$ **b.** $P(A' \mid B)$
c. $P(A \mid B')$ **d.** $P(B \mid A)$

2.5.10 A computer system uses passwords that contain exactly eight characters, and each character is one of the 26 lowercase letters (a–z) or 26 uppercase letters (A–Z) or 10 integers (0–9). Let Ω denote the set of all possible passwords. Suppose that all passwords in Ω are equally likely. Determine the probability for each of the following:

a. Password contains all lowercase letters given that it contains only letters
b. Password contains at least 1 uppercase letter given that it contains only letters
c. Password contains only even numbers given that it contains all numbers

2.5.11 Suppose that a patient is selected randomly from those described in Exercise 2.4.11. Let A denote the event that the patient is in group 1, and let B denote the event that there is no progression. Determine the following probabilities:

a. $P(B \mid A)$ **b.** $P(A \mid B)$
c. $P(A \mid B')$ **d.** $P(A' \mid B)$

Exercises for Section 2.6

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

2.6.1 WP Suppose that $P(A \mid B) = 0.4$ and $P(B) = 0.5$. Determine the following:

a. $P(A \cap B)$
b. $P(A' \cap B)$

2.6.2 WP Suppose that $P(A \mid B) = 0.2$, $P(A \mid B') = 0.3$, and $P(B) = 0.8$. What is $P(A)$?

2.6.3 WP The probability is 1% that an electrical connector that is kept dry fails during the warranty period. If the connector is ever wet, the probability of a failure during the warranty period is 5%. If 90% of the connectors are kept dry and 10% are wet, what proportion of connectors fail during the warranty period?

2.6.4 WP Heart failures are due to either natural occurrences (87%) or outside factors (13%). Outside factors are related to induced substances (73%) or foreign objects (27%). Natural occurrences are caused by arterial blockage (56%), disease (27%), and infection (e.g., staph infection) (17%).

a. Determine the probability that a failure is due to an induced substance.
b. Determine the probability that a failure is due to disease or infection.

2.6.5 WP SS The edge roughness of slit paper products increases as knife blades wear. Only 1% of products slit with new blades have rough edges, 3% of products slit with blades of average sharpness exhibit roughness, and 5% of products slit with worn blades exhibit roughness. If 25% of the blades in manufacturing are new, 60% are of average sharpness, and 15% are worn, what is the proportion of products that exhibit edge roughness?

2.6.6 WP A lot of 100 semiconductor chips contains 20 that are defective.

a. Two are selected, at random, without replacement, from the lot. Determine the probability that the second chip selected is defective.
b. Three are selected, at random, without replacement, from the lot. Determine the probability that all are defective.

2.6.7 WP Computer keyboard failures are due to faulty electrical connects (12%) or mechanical defects (88%). Mechanical defects are related to loose keys (27%) or improper assembly (73%). Electrical connect defects are caused by defective wires (35%), improper connections (13%), or poorly welded wires (52%).

a. Find the probability that a failure is due to loose keys.
b. Find the probability that a failure is due to improperly connected or poorly welded wires.

2.6.8 An article in the *British Medical Journal* ["Comparison of Treatment of Renal Calculi by Operative Surgery, Percutaneous Nephrolithotomy, and Extracorporeal Shock Wave Lithotripsy" (1986, Vol. 82, pp. 879–892)] provided the following discussion of success rates in kidney stone removals. Open surgery had a success rate of 78% (273/350) and a newer method, percutaneous nephrolithotomy (PN), had a success rate of 83% (289/350). This newer method looked better, but the results changed when stone diameter was considered. For stones with diameters less than 2 centimeters, 93% (81/87) of cases of open surgery were successful compared with only 83% (234/270) of cases of PN. For stones greater than or equal to 2 centimeters, the success rates were 73% (192/263) and 69% (55/80) for open surgery and PN, respectively. Open surgery is better for both stone sizes, but less successful in total. In 1951, E. H. Simpson pointed out this apparent contradiction (known as **Simpson's paradox**), and the hazard still persists today. Explain how open surgery can be better for both stone sizes but worse in total.

2.6.9 SS Consider the hospital emergency room data in Example 2.6. Let A denote the event that a visit is to hospital 4 and let B denote the event that a visit results in LWBS (at any hospital). Determine the following probabilities.

a. $P(A \cap B)$ **b.** $P(A \cup B)$ **c.** $P'(A' \cup B')$

d. Use the total probability rule to determine $P(A)$

2.6.10 Suppose that a patient is selected randomly from those described in Exercise 2.4.11. Let A denote the event that the patient is in group 1, and let B denote the event for which there is no progression. Determine the following probabilities:

a. $P(A \cap B)$ **b.** $P(B)$ **c.** $P(A' \cap B)$

d. $P(A \cup B)$ **e.** $P(A' \cup B)$

2.6.11 A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. A specific design is randomly generated by the Web server when you visit the site. Determine the probability that the ad color is red and the font size is not the smallest one.

2.6.12 Similar to the hospital schedule in Example 2.9, suppose that an operating room needs to schedule three knee, four hip, and five shoulder surgeries. Assume that all schedules are equally likely. Determine the following probabilities:

a. All hip surgeries are completed first given that all knee surgeries are last.

b. The schedule begins with a hip surgery given that all knee surgeries are last.

c. The first and last surgeries are hip surgeries given that knee surgeries are scheduled in time periods 2 through 4.

d. The first two surgeries are hip surgeries given that all knee surgeries are last.

Exercises for Section 2.7

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

2.7.1 WP If $P(A \mid B) = 0.3$, $P(B) = 0.8$, and $P(A) = 0.3$, are the events B and the complement of A independent?

2.7.2 WP If $P(A) = 0.2$, $P(B) = 0.2$, and A and B are mutually exclusive, are they independent?

2.7.3 WP A batch of 500 containers of frozen orange juice contains 5 that are defective. Two are selected, at random, without replacement, from the batch. Let A and B denote the events that the first and second containers selected are defective, respectively.

a. Are A and B independent events?

b. If the sampling were done with replacement, would A and B be independent?

2.7.4 WP Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

		Shock Resistance	
		High	Low
Scratch Resistance	High	70	9
	Low	16	5

Let A denote the event that a disk has high shock resistance, and let B denote the event that a disk has high scratch resistance. Are events A and B independent?

2.7.5 WP SS Redundant array of inexpensive disks (RAID) is a technology that uses multiple hard drives to increase the speed of data transfer and provide instant data backup. Suppose that the probability of any hard drive failing in a day is 0.001 and the drive failures are independent.

a. A RAID 0 scheme uses two hard drives, each containing a mirror image of the other. What is the probability of data loss? Assume that data loss occurs if both drives fail within the same day.

b. A RAID 1 scheme splits the data over two hard drives. What is the probability of data loss? Assume that data loss occurs if at least one drive fails within the same day.

2.7.6 WP A test of a printed circuit board uses a random test pattern with an array of 10 bits and each is equally likely to be 0 or 1. Assume the bits are independent.

a. What is the probability that all bits are 1s?

b. What is the probability that all bits are 0s?

c. What is the probability that exactly 5 bits are 1s and 5 bits are 0s?

2.7.7 WP The probability that a lab specimen contains high levels of contamination is 0.10. Five samples are checked, and the samples are independent.

a. What is the probability that none contain high levels of contamination?

b. What is the probability that exactly one contains high levels of contamination?

c. What is the probability that at least one contains high levels of contamination?

2.7.8 WP A player of a video game is confronted with a series of four opponents and an 80% probability of defeating each opponent. Assume that the results from opponents are independent (and that when the player is defeated by an opponent the game ends).

a. What is the probability that a player defeats all four opponents in a game?

b. What is the probability that a player defeats at least two opponents in a game?

c. If the game is played three times, what is the probability that the player defeats all four opponents at least once?

2.7.9 WP Eight cavities in an injection-molding tool produce plastic connectors that fall into a common stream. A sample is chosen every several minutes. Assume that the samples are independent.

a. What is the probability that five successive samples were all produced in cavity 1 of the mold?

b. What is the probability that five successive samples were all produced in the same cavity of the mold?

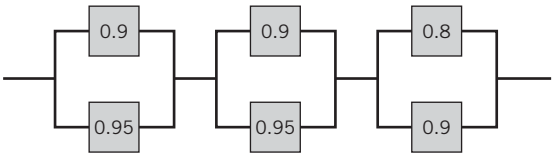
c. What is the probability that four out of five successive samples were produced in cavity 1 of the mold?

2.7.10 A credit card contains 16 digits. It also contains the month and year of expiration. Suppose there are 1 million credit card holders with unique card numbers. A hacker randomly selects a 16-digit credit card number.

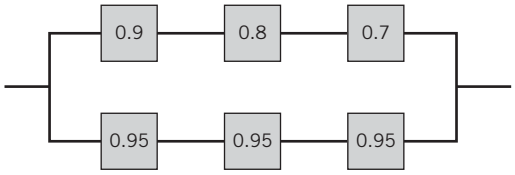
a. What is the probability that it belongs to a user?

b. Suppose a hacker has a 25% chance of correctly guessing the year your card expires and randomly selects 1 of the 12 months. What is the probability that the hacker correctly selects the month and year of expiration?

2.7.11 SS The following circuit operates if and only if there is a path of functional devices from left to right. The probability that each device functions is as shown. Assume that the probability that a device is functional does not depend on whether or not other devices are functional. What is the probability that the circuit operates?



2.7.12 VS The following circuit operates if and only if there is a path of functional devices from left to right. The probability that each device functions is as shown. Assume that the probability that a device is functional does not depend on whether or not other devices are functional. What is the probability that the circuit operates?



2.7.13 Consider the hospital emergency room data in Example 2.6. Let A denote the event that a visit is to hospital 4, and let B denote the event that a visit results in LWBS (at any hospital). Are these events independent?

2.7.14 A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. A specific design is randomly generated by the Web server when you visit the site. Let A denote the event that the design color is red, and let B denote the event that the font size is not the smallest one. Are A and B independent events? Explain why or why not.

2.7.15 An integrated circuit contains 10 million logic gates (each can be a logical AND or OR circuit). Assume the probability of a gate failure is p and that the failures are independent. The integrated circuit fails to function if any gate fails. Determine the value for p so that the probability that the integrated circuit functions is 0.95.

2.7.16 Table 2.1 provides data on wafers categorized by location and contamination levels. Let A denote the event that contamination is *low*, and let B denote the event that the location is *center*. Are A and B independent? Why or why not?

Exercises for Section 2.8

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

2.8.1 WP SS VS Customers are used to evaluate preliminary product designs. In the past, 95% of highly successful products received good reviews, 60% of moderately successful products received good reviews, and 10% of poor products received good reviews. In addition, 40% of products have been highly successful, 35% have been moderately successful, and 25% have been poor products.

a. What is the probability that a product attains a good review?

b. If a new design attains a good review, what is the probability that it will be a highly successful product?

c. If a product does not attain a good review, what is the probability that it will be a highly successful product?

2.8.2 WP Suppose that $P(A \mid B) = 0.7$, $P(A) = 0.5$, and $P(B) = 0.2$. Determine $P(B \mid A)$.

2.8.3 WP A new analytical method to detect pollutants in water is being tested. This new method of chemical analysis is important because, if adopted, it could be used to detect three different contaminants—organic pollutants, volatile solvents, and chlorinated compounds—instead of having to use a single test for each pollutant. The makers of the test claim that it can detect high levels of organic pollutants with 99.7% accuracy, volatile solvents with 99.95% accuracy, and chlorinated compounds with 89.7% accuracy. If a pollutant is not present, the test does not signal. Samples are prepared for the calibration of the test and 60% of them are contaminated with organic pollutants, 27% with volatile solvents, and 13% with traces of chlorinated compounds. A test sample is selected randomly.

a. What is the probability that the test will signal?

b. If the test signals, what is the probability that chlorinated compounds are present?

2.8.4 WP Software to detect fraud in consumer phone cards tracks the number of metropolitan areas where calls originate each day. It is found that 1% of the legitimate users originate calls from two or more metropolitan areas in a single day. However, 30% of fraudulent users originate calls from two or more metropolitan areas in a single day. The proportion of fraudulent users is 0.01%. If the same user originates calls from two or more metropolitan areas in a single day, what is the probability that the user is fraudulent?

2.8.5 Consider the hospital emergency room data in Example 2.6. Use Bayes' theorem to calculate the probability that a person visits its hospital 4 given they are LWBS.

2.8.6 Suppose that a patient is selected randomly from those described in Exercise 2.4.11. Let A denote the event that the patient is in group 1, and let B denote the event that there is no progression. Determine the following probabilities:

a. $P(B)$ **b.** $P(B \mid A)$ **c.** $P(A \mid B)$

2.8.7 SS Two Web colors are used for a site advertisement. If a site visitor arrives from an affiliate, the probabilities of the blue or green colors being used in the advertisement are 0.8 and 0.2, respectively. If the site visitor arrives from a search site, the probabilities of blue and green colors in the advertisement are 0.4 and 0.6, respectively. The proportions of visitors from affiliates and search sites are 0.3 and 0.7, respectively. What is the probability that a visitor is from a search site given that the blue ad was viewed?

2.8.8 A recreational equipment supplier finds that among orders that include tents, 40% also include sleeping mats. Only 5% of orders that do not include tents do include sleeping mats. Also, 20% of orders include tents. Determine the following probabilities:

a. The order includes sleeping mats.
b. The order includes a tent given it includes sleeping mats.

2.8.9 An e-mail filter is planned to separate valid e-mails from spam. The word *free* occurs in 60% of the spam messages and only 4% of the valid messages. Also, 20% of the messages are spam. Determine the following probabilities:

a. The message contains *free*.
b. The message is spam given that it contains *free*.
c. The message is valid given that it does not contain *free*.

Exercises for Section 2.9

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

2.9.1 SS Decide whether a discrete or continuous random variable is the best model for each of the following variables:

a. The number of cracks exceeding one-half inch in 10 miles of an interstate highway.
b. The weight of an injection-molded plastic part.
c. The number of molecules in a sample of gas.
d. The concentration of output from a reactor.
e. The current in an electronic circuit.

2.9.2 Decide whether a discrete or continuous random variable is the best model for each of the following variables:

a. The time until a projectile returns to earth.
b. The number of times a transistor in a computer memory changes state in one operation.
c. The volume of gasoline that is lost to evaporation during the filling of a gas tank.
d. The outside diameter of a machined shaft.

2.9.3 Decide whether a discrete or continuous random variable is the best model for each of the following variables:

a. The time for a computer algorithm to assign an image to a category.
b. The number of bytes used to store a file in a computer.
c. The ozone concentration in micrograms per cubic meter.
d. The ejection fraction (volumetric fraction of blood pumped from a heart ventricle with each beat).
e. The fluid flow rate in liters per minute.

Supplemental Exercises for Chapter 2

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

2.S4 WP Samples of laboratory glass are in small, light packaging or heavy, large packaging. Suppose that 2% and 1%, respectively, of the sample shipped in small and large packages, respectively, break during transit. If 60% of the samples are shipped in large packages and 40% are shipped in small packages, what proportion of samples break during shipment?

2.S5 WP SS Samples of a cast aluminum part are classified on the basis of surface finish (in microinches) and edge finish. The results of 100 parts are summarized as follows:

		Edge Finish	
		Excellent	Good
Surface finish	Excellent	80	2
	Good	10	8

Let A denote the event that a sample has excellent surface finish, and let B denote the event that a sample has excellent edge finish. If a part is selected at random, determine the following probabilities:

a. $P(A)$ **b.** $P(B)$ **c.** $P(A')$
d. $P(A \cap B)$ **e.** $P(A \cup B)$ **f.** $P(A' \cup B)$

2.S6 WP A lot contains 15 castings from a local supplier and 25 castings from a supplier in the next state. Two castings are

selected randomly, without replacement, from the lot of 40. Let A be the event that the first casting selected is from the local supplier, and let B denote the event that the second casting is selected from the local supplier. Determine:

a. $P(A)$ **b.** $P(B \mid A)$ **c.** $P(A \cap B)$ **d.** $P(A \cup B)$

Suppose that 3 castings are selected at random, without replacement, from the lot of 40. In addition to the definitions of events A and B, let C denote the event that the third casting selected is from the local supplier. Determine:

e. $P(A \cap B \cap C)$ **f.** $P(A \cap B \cap C')$

2.S7 If A, B, and C are mutually exclusive events, is it possible for $P(A) = 0.3$, $P(B) = 0.4$, and $P(C) = 0.5$? Why or why not?

2.S8 WP Incoming calls to a customer service center are classified as complaints (75% of calls) or requests for information (25% of calls). Of the complaints, 40% deal with computer equipment that does not respond and 57% deal with incomplete software installation; in the remaining 3% of complaints, the user has improperly followed the installation instructions. The requests for information are evenly divided on technical questions (50%) and requests to purchase more products (50%).

a. What is the probability that an incoming call to the customer service center will be from a customer who has not followed installation instructions properly?

b. Find the probability that an incoming call is a request for purchasing more products.

2.S9 WP In the manufacturing of a chemical adhesive, 3% of all batches have raw materials from two different lots. This occurs when holding tanks are replenished and the remaining portion of a lot is insufficient to fill the tanks.

Only 5% of batches with material from a single lot require reprocessing. However, the viscosity of batches consisting of two or more lots of material is more difficult to control, and 40% of such batches require additional processing to achieve the required viscosity.

Let A denote the event that a batch is formed from two different lots, and let B denote the event that a lot requires additional processing. Determine the following probabilities:

a. $P(A)$ **b.** $P(A')$ **c.** $P(B \mid A)$
d. $P(B \mid A')$ **e.** $P(A \cap B)$ **f.** $P(A \cap B')$ **g.** $P(B)$

2.S10 Semiconductor lasers used in optical storage products require higher power levels for write operations than for read operations. High-power-level operations lower the useful life of the laser.

Lasers in products used for backup of higher-speed magnetic disks primarily write, and the probability that the useful life exceeds five years is 0.95. Lasers that are in products that are used for main storage spend approximately an equal amount of time reading and writing, and the probability that the useful life exceeds five years is 0.995. Now, 25% of the products from a manufacturer are used for backup and 75% of the products are used for main storage.

Let A denote the event that a laser's useful life exceeds five years, and let B denote the event that a laser is in a product that is used for backup.

Use a tree diagram to determine the following:

a. $P(B)$ **b.** $P(A \mid B)$ **c.** $P(A \mid B')$
d. $P(A \cap B)$ **e.** $P(A \cap B')$ **f.** $P(A)$

g. What is the probability that the useful life of a laser exceeds five years?

h. What is the probability that a laser that failed before five years came from a product used for backup?

2.S11 WP SS A congested computer network has a 0.002 probability of losing a data packet, and packet losses are independent events. A lost packet must be resent.

a. What is the probability that an e-mail message with 100 packets will need to be resent?

b. What is the probability that an e-mail message with 3 packets will need exactly 1 to be resent?

c. If 10 e-mail messages are sent, each with 100 packets, what is the probability that at least 1 message will need some packets to be resent?

2.S12 WP GO Tutorial An electronic storage device uses an error recovery procedure that requires an immediate satisfactory readback of any written data. If the readback is not successful after three writing operations, that sector of the device is eliminated as unacceptable for data storage. On an acceptable portion of the device, the probability of a satisfactory readback is 0.98. Assume the readbacks are independent. What is the probability that an acceptable portion of the device is eliminated as unacceptable for data storage?

2.S13 Energy released from cells breaks the molecular bond and converts ATP (adenosine triphosphate) into ADP (adenosine diphosphate). Storage of ATP in muscle cells (even for an athlete) can sustain maximal muscle power only for less than five seconds (a short dash). Three systems are used to replenish ATP—phosphagen system, glycogen-lactic acid system (anaerobic), and aerobic respiration—but the first is useful only for less than 10 seconds, and even the second system provides less than two minutes of ATP. An endurance athlete needs to perform below the anaerobic threshold to sustain energy for extended periods. A sample of 100 individuals is described by the energy system used in exercise at different intensity levels.

	Primarily Aerobic	
Period	Yes	No
1	50	7
2	13	30

Let A denote the event that an individual is in period 2, and let B denote the event that the energy is primarily aerobic. Determine the number of individuals in

a. $A' \cap B$ **b.** B' **c.** $A \cup B$

2.S14 WP The probability that a customer's order is not shipped on time is 0.05. A particular customer places three orders, and the orders are placed far enough apart in time that they can be considered to be independent events.

a. What is the probability that all are shipped on time?

b. What is the probability that exactly one is not shipped on time?

c. What is the probability that two or more orders are not shipped on time?

2.S15 In circuit testing of printed circuit boards, each board either fails or does not fail the test. A board that fails the test is then checked further to determine which one of five defect types is the primary failure mode. Represent the sample space for this experiment.

2.S16 WP Transactions to a computer database are either new items or changes to previous items. The addition of an item can be completed in less than 100 milliseconds 90% of the time, but only 20% of changes to a previous item can be completed in less than this time. If 30% of transactions are changes, what is the probability that a transaction can be completed in less than 100 milliseconds?

2.S17 Let E_1, E_2, and E_3 denote the samples that conform to a percentage of solids specification, a molecular weight specification, and a color specification, respectively. A total of 240 samples are classified by the E_1, E_2, and E_3 specifications, where *yes* indicates that the sample conforms.

E_3 yes

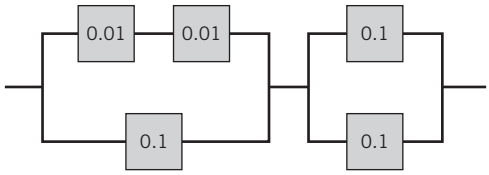
		E_2		
		Yes	No	Total
E_1	Yes	200	1	201
	No	5	4	9
Total		205	5	210

E_3 no

		E_2		
		Yes	No	Total
E_1	Yes	20	4	24
	No	6	0	6
Total		26	4	30

a. Are E_1, E_2, and E_3 mutually exclusive events?
b. Are E_1', E_2', and E_3' mutually exclusive events?
c. What is $P(E_1'$ or E_2' or $E_3')$?
d. What is the probability that a sample conforms to all three specifications?
e. What is the probability that a sample conforms to the E_1 or E_3 specification?
f. What is the probability that a sample conforms to the E_1 or E_2 or E_3 specification?

2.S18 The following circuit operates if and only if there is a path of functional devices from left to right. Assume devices fail independently and that the probability of *failure* of each device is as shown. What is the probability that the circuit operates?



2.S19 WP A steel plate contains 20 bolts. Assume that 5 bolts are not torqued to the proper limit. Four bolts are selected at random, without replacement, to be checked for torque.

a. What is the probability that all 4 of the selected bolts are torqued to the proper limit?
b. What is the probability that at least 1 of the selected bolts is *not* torqued to the proper limit?

2.S20 The British government has stepped up its information campaign regarding foot-and-mouth disease by mailing brochures to farmers around the country. It is estimated that 99% of Scottish farmers who receive the brochure possess enough information to deal with an outbreak of the disease, but only 90% of those without the brochure can deal with an outbreak. After the first three months of mailing, 95% of the farmers in Scotland had received the informative brochure. Compute the probability that a randomly selected farmer will have enough information to deal effectively with an outbreak of the disease.

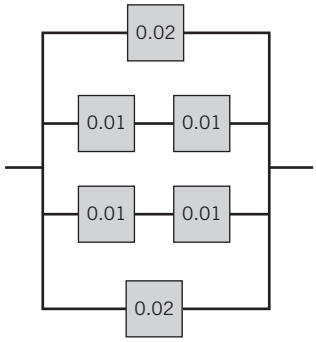
2.S21 WP SS It is known that two defective cellular phones were erroneously sent to a shipping lot that now has a total of 75 phones. A sample of phones will be selected from the lot without replacement.

a. If three phones are inspected, determine the probability that exactly one of the defective phones will be found.
b. If three phones are inspected, determine the probability that both defective phones will be found.
c. If 73 phones are inspected, determine the probability that both defective phones will be found. (*Hint*: Work with the phones that remain in the lot.)

2.S22 WP An encryption–decryption system consists of three elements: encode, transmit, and decode. A faulty encode occurs in 0.5% of the messages processed, transmission errors occur in 1% of the messages, and a decode error occurs in 0.1% of the messages. Assume the errors are independent.

a. What is the probability of a completely defect-free message?
b. What is the probability of a message that has either an encode or a decode error?

2.S23 The following circuit operates if and only if there is a path of functional devices from left to right. Assume that devices fail independently and that the probability of *failure* of each device is as shown. What is the probability that the circuit does not operate?



2.S24 WP A robotic insertion tool contains 10 primary components. The probability that any component fails during the warranty period is 0.01. Assume that the components fail independently and that the tool fails if any component fails. What is the probability that the tool fails during the warranty period?

2.S25 A machine tool is idle 15% of the time. You request immediate use of the tool on five different occasions during the year. Assume that your requests represent independent events.

a. What is the probability that the tool is idle at the time of all of your requests?

b. What is the probability that the machine is idle at the time of exactly four of your requests?

c. What is the probability that the tool is idle at the time of at least three of your requests?

2.S26 A company that tracks the use of its Web site determined that the more pages a visitor views, the more likely the visitor is to request more information (RMI). Use the following table to answer the questions:

Number of pages viewed:	1	2	3	4 or more
Percentage of visitors:	40	30	20	10
Percentage who RMI	10	10	20	40

a. What is the probability that a visitor to the Web site RMIs?

b. If a visitor RMIs, what is the probability that the visitor viewed four or more pages?

2.S27 SS An article in *Genome Research* ["An Assessment of Gene Prediction Accuracy in Large DNA Sequences" (2000, Vol. 10, pp. 1631–1642)] considered the accuracy of commercial software to predict nucleotides in gene sequences. The following table shows the number of sequences for which the programs produced predictions and the number of nucleotides correctly predicted (computed globally from the total number of prediction successes and failures on all sequences).

	Number of Sequences	Proportion
GenScan	177	0.93
Blastx default	175	0.91
Blastx topcomboN	174	0.97
Blastx 2 stages	175	0.90
GeneWise	177	0.98
Procrustes	177	0.93

Assume the prediction successes and failures are independent among the programs.

a. What is the probability that all programs predict a nucleotide correctly?

b. What is the probability that all programs predict a nucleotide incorrectly?

c. What is the probability that at least one Blastx program predicts a nucleotide correctly?

2.S28 WP A batch contains 36 bacteria cells. Assume that 12 of the cells are not capable of cellular replication. Of the cells, 6 are selected at random, without replacement, to be checked for replication.

a. What is the probability that all 6 of the selected cells are able to replicate?

b. What is the probability that at least 1 of the selected cells is not capable of replication?

2.S29 A computer system uses passwords that are exactly seven characters, and each character is one of the 26 letters (a–z) or 10 integers (0–9). Uppercase letters are not used.

a. How many passwords are possible?

b. If a password consists of exactly 6 letters and 1 number, how many passwords are possible?

c. If a password consists of 5 letters followed by 2 numbers, how many passwords are possible?

2.S30 Consider the treatments in Exercise 2.1.25. Suppose a patient is selected randomly. Let A denote the event that the patient is treated with ribavirin plus interferon alfa or interferon alfa, and let B denote the event that the response is complete. Determine the following probabilities.

a. $P(A \mid B)$ **b.** $P(B \mid A)$ **c.** $P(A \cap B)$ **d.** $P(A \cup B)$

2.S31 Consider the patient groups in Exercise 2.4.11. Suppose a patient is selected randomly. Let A denote the event that the patient is in group 1 or 2, and let B denote the event that there is no progression. Determine the following probabilities:

a. $P(A \mid B)$ **b.** $P(B \mid A)$ **c.** $P(A \cap B)$ **d.** $P(A \cup B)$

Chapter 3 Exercises

Exercises for Section 3.1

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

For each of the following exercises, determine the range (possible values) of the random variable.

3.1.1 WP The random variable is the number of nonconforming solder connections on a printed circuit board with 1000 connections.

3.1.2 WP A batch of 500 machined parts contains 10 that do not conform to customer requirements. The random variable is the number of parts in a sample of five parts that do not conform to customer requirements.

3.1.3 WP A batch of 500 machined parts contains 10 that do not conform to customer requirements. Parts are selected successively, without replacement, until a nonconforming part is obtained. The random variable is the number of parts selected.

3.1.4 The random variable is the number of computer clock cycles required to complete a selected arithmetic calculation.

3.1.5 The random variable is the number of surface flaws in a large coil of galvanized steel.

3.1.6 WP A group of 10,000 people are tested for a gene called *Ifi202* that has been found to increase the risk for lupus. The random variable is the number of people who carry the gene.

3.1.7 WP SS The number of mutations in a nucleotide sequence of length 40,000 in a DNA strand after exposure to radiation is measured. Each nucleotide may be mutated.

3.1.8 A healthcare provider schedules 30 minutes for each patient's visit, but some visits require extra time. The random variable is the number of patients treated in an eight-hour day.

3.1.9 The sample space of a random experiment is $\{a, b, c, d, e, f\}$, and each outcome is equally likely. A random variable is defined as follows:

outcome	a	b	c	d	e	f
x	0	0	1.5	1.5	2	3

Determine the probability mass function of X. Use the probability mass function to determine the following probabilities:

a. $P(X = 1.5)$ **b.** $P(0.5 < X < 2.7)$
c. $P(X > 3)$ **d.** $P(0 \le X < 2)$
e. $P(X = 0 \text{ or } X = 2)$

For Exercises 3.1.10 to 3.1.12, verify that the following functions are probability mass functions, and determine the requested probabilities.

3.1.10 WP $f(x) = (8/7)(1/2)^x,\ x = 1, 2, 3$

a. $P(X \le 1)$ **b.** $P(X > 1)$
c. $P(2 < X < 6)$ **d.** $P(X \le 1 \text{ or } X > 1)$

3.1.11 WP SS VS $f(x) = \dfrac{2x+1}{25},\quad x = 0, 1, 2, 3, 4$

a. $P(X = 4)$ **b.** $P(X \le 1)$
c. $P(2 \le X < 4)$ **d.** $P(X > -10)$

3.1.12 Consider the hospital patients in Example 2.6. Two patients are selected randomly, with replacement, from the total patients at Hospital 1. What is the probability mass function of the number of patients in the sample who are admitted?

3.1.13

x	1.25	1.5	1.75	2	2.25
$f(x)$	0.2	0.4	0.1	0.2	0.1

a. $P(X \ge 2)$ **b.** $P(X < 1.65)$
c. $P(X = 1.5)$ **d.** $P(X < 1.3 \text{ or } X > 2.1)$

3.1.14 WP An article in *Knee Surgery, Sports Traumatology, Arthroscopy* ["Arthroscopic Meniscal Repair with an Absorbable Screw: Results and Surgical Technique" (2005, Vol. 13, pp. 273–279)] cited a success rate of more than 90% for meniscal tears with a rim width under 3 mm, but only a 67% success rate for tears of 3–6 mm. If you are unlucky enough to suffer a meniscal tear of under 3 mm on your left knee and one of width 3–6 mm on your right knee, what is the probability mass function of the number of successful surgeries? Assume that the surgeries are independent.

3.1.15 WP In a semiconductor manufacturing process, three wafers from a lot are tested. Each wafer is classified as *pass* or *fail*. Assume that the probability that a wafer passes the test is 0.8 and that wafers are independent. Determine the probability mass function of the number of wafers from a lot that pass the test.

3.1.16 WP Marketing estimates that a new instrument for the analysis of soil samples will be very successful, moderately successful, or unsuccessful with probabilities 0.3, 0.6, and 0.1, respectively. The yearly revenue associated with a very successful, moderately successful, or unsuccessful product is \$10 million, \$5 million, and \$1 million, respectively. Let the random variable X denote the yearly revenue of the product. Determine the probability mass function of X.

3.1.17 WP SS An assembly consists of three mechanical components. Suppose that the probabilities that the first, second, and third components meet specifications are 0.95, 0.98, and 0.99, respectively. Assume that the components are independent. Determine the probability mass function of the number of components in the assembly that meet specifications.

3.1.18 The distribution of the time until a Web site changes is important to Web crawlers that search engines use to maintain current information about Web sites. The distribution of the time until change (in days) of a Web site is approximated in the following table.

Days until Changes	Probability
1.5	0.05
3.0	0.25
4.5	0.35
5.0	0.20
7.0	0.15

Calculate the probability mass function of the days until change.

3.1.19 WP Actual lengths of stay at a hospital's emergency department in 2009 are shown in the following table (rounded to the nearest hour). Length of stay is the total of wait and service times. Some longer stays are also approximated as 15 hours in this table.

Hours	Count	Percent
1	19	3.80
2	51	10.20
3	86	17.20
4	102	20.40
5	87	17.40
6	62	12.40
7	40	8.00
8	18	3.60
9	14	2.80
10	11	2.20
15	10	2.00

Calculate the probability mass function of the wait time for service.

Exercises for Section 3.2

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

3.2.1 Determine the cumulative distribution function for the random variable in Exercise 3.1.9.

3.2.2 WP Determine the cumulative distribution function of the random variable in Exercise 3.1.10.

3.2.3 Determine the cumulative distribution function for the random variable in Exercise 3.1.11.

3.2.4 WP Determine the cumulative distribution function for the random variable in Exercise 3.1.12; also determine the following probabilities:

a. $P(X < 1.5)$ **b.** $P(X \le 3)$

c. $P(X > 2)$ **d.** $P(1 < X \le 2)$

3.2.5 WP SS Determine the cumulative distribution function for the random variable in Exercise 3.1.13.

3.2.6 Determine the cumulative distribution function for the random variable in Exercise 3.1.16.

Verify that the following functions are cumulative distribution functions, and determine the probability mass function and the requested probabilities.

3.2.7 WP $F(x) = \begin{cases} 0 & x < 1 \\ 0.5 & 1 \le x < 3 \\ 1 & 3 \le x \end{cases}$

a. $P(X \le 3)$ **b.** $P(X \le 2)$

c. $P(1 \le X \le 2)$ **d.** $P(X > 2)$

3.2.8 WP $F(x) = \begin{cases} 0 & x < -10 \\ 0.25 & -10 \le x < 30 \\ 0.75 & 30 \le x < 50 \\ 1 & 50 \le x \end{cases}$

a. $P(X \le 50)$ **b.** $P(X \le 40)$

c. $P(40 \le X \le 60)$ **d.** $P(X < 0)$

e. $P(0 \le X < 10)$ **f.** $P(-10 < X < 10)$

Exercises for Section 3.3

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

3.3.1 WP If the range of X is the set $\{0, 1, 2, 3, 4\}$ and $P(X = x) = 0.2$, determine the mean and variance of the random variable.

3.3.2 WP Determine the mean and variance of the random variable in Exercise 3.1.10.

3.3.3 WP VS Determine the mean and variance of the random variable in Exercise 3.1.11.

3.3.4 WP Determine the mean and variance of the random variable in Exercise 3.1.12.

3.3.5 WP Determine the mean and variance of the random variable in Exercise 3.1.13.

3.3.6 In a NiCd battery, a fully charged cell is composed of nickelic hydroxide. Nickel is an element that has multiple oxidation states. Assume the following proportions of the states:

Nickel Charge	Proportions Found
0	0.17
+2	0.35
+3	0.33
+4	0.15

a. Determine the cumulative distribution function of the nickel charge.

b. Determine the mean and variance of the nickel charge.

3.3.7 WP SS The range of the random variable X is [0, 1, 2, 3, x] where x is unknown. If each value is equally likely and the mean of X is 6, determine x.

3.3.8 WP Trees are subjected to different levels of carbon dioxide atmosphere with 6% of them in a minimal growth condition at 350 parts per million (ppm), 10% at 450 ppm (slow growth), 47% at 550 ppm (moderate growth), and 37% at 650 ppm (rapid growth). What are the mean and standard deviation of the carbon dioxide atmosphere (in ppm) for these trees in ppm?

Exercises for Section 3.4

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

3.4.1 WP Assume that the wavelengths of photosynthetically active radiations (PAR) are uniformly distributed at integer nanometers in the red spectrum from 675 to 700 nm.

a. What are the mean and variance of the wavelength distribution for this radiation?

b. If the wavelengths are uniformly distributed at integer nanometers from 75 to 100 nanometers, how do the mean and variance of the wavelength distribution compare to the previous part? Explain.

3.4.2 Let the random variable X have a discrete uniform distribution on the integers $0 \le x \le 99$. Determine the mean and variance of X.

3.4.3 WP SS Suppose that X has a discrete uniform distribution on the integers 0 through 9. Determine the mean, variance, and standard deviation of the random variable $Y = 5X$ and compare to the corresponding results for X.

3.4.4 WP Thickness measurements of a coating process are made to the nearest hundredth of a millimeter. The thickness measurements are uniformly distributed with values 0.15, 0.16, 0.17, 0.18, and 0.19. Determine the mean and variance of the coating thickness for this process.

3.4.5 Suppose that 1000 seven-digit telephone numbers within your area code are dialed randomly. What is the probability that your number is called?

3.4.6 Each multiple-choice question on an exam has four choices. Suppose that there are 10 questions and the choice is selected randomly and independently for each question. Let X denote the number of questions answered correctly. Does X have a discrete uniform distribution? Why or why not?

Exercises for Section 3.5

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

3.5.1 For each scenario (a)–(j), state whether or not the binomial distribution is a reasonable model for the random variable and why. State any assumptions you make.

a. A production process produces thousands of temperature transducers. Let X denote the number of nonconforming transducers in a sample of size 30 selected at random from the process.

b. From a batch of 50 temperature transducers, a sample of size 30 is selected without replacement. Let X denote the number of nonconforming transducers in the sample.

c. Four identical electronic components are wired to a controller that can switch from a failed component to one of the remaining spares. Let X denote the number of components that have failed after a specified period of operation.

d. Let X denote the number of accidents that occur along the federal highways in Arizona during a one-month period.

e. Let X denote the number of correct answers by a student taking a multiple-choice exam in which a student can eliminate some of the choices as being incorrect in some questions and all of the incorrect choices in other questions.

f. Defects occur randomly over the surface of a semiconductor chip. However, only 80% of defects can be found by testing. A sample of 40 chips with one defect each is tested. Let X denote the number of chips in which the test finds a defect.

g. Errors in a digital communication channel occur in bursts that affect several consecutive bits. Let X denote the number of bits in error in a transmission of 100,000 bits.

h. Let X denote the number of surface flaws in a large coil of galvanized steel.

3.5.2 WP Let X be a binomial random variable with $p = 0.1$ and $n = 10$. Calculate the following probabilities.

a. $P(X \le 2)$ **b.** $P(X > 8)$
c. $P(X = 4)$ **d.** $P(5 \le X \le 7)$

3.5.3 WP SS The random variable X has a binomial distribution with $n = 10$ and $p = 0.01$. Determine the following probabilities.

a. $P(X = 5)$ **b.** $P(X \le 2)$
c. $P(X \ge 9)$ **d.** $P(3 \le X < 5)$

3.5.4 The random variable X has a binomial distribution with $n = 10$ and $p = 0.5$. Sketch the probability mass function of X.

a. What value of X is most likely?

b. What value(s) of X is(are) least likely?

c. Repeat the previous parts with $p = 0.01$.

3.5.5 Determine the cumulative distribution function of a binomial random variable with $n = 3$ and $p = 1/4$.

3.5.6 WP An electronic product contains 40 integrated circuits. The probability that any integrated circuit is defective is 0.01, and the integrated circuits are independent. The product operates only if there are no defective integrated circuits. What is the probability that the product operates?

3.5.7 SS VS The phone lines to an airline reservation system are occupied 40% of the time. Assume that the events that the lines are occupied on successive calls are independent. Assume that 10 calls are placed to the airline.

a. What is the probability that for exactly three calls, the lines are occupied?

b. What is the probability that for at least one call, the lines are not occupied?

c. What is the expected number of calls in which the lines are all occupied?

3.5.8 A multiple-choice test contains 25 questions, each with four answers. Assume that a student just guesses on each question.

a. What is the probability that the student answers more than 20 questions correctly?

b. What is the probability that the student answers fewer than 5 questions correctly?

3.5.9 WP Samples of rejuvenated mitochondria are mutated (defective) in 1% of cases. Suppose that 15 samples are studied and can be considered to be independent for mutation. Determine the following probabilities.

a. No samples are mutated.

b. At most one sample is mutated.

c. More than half the samples are mutated.

3.5.10 WP Heart failure is due to either natural occurrences (87%) or outside factors (13%). Outside factors are related to induced substances or foreign objects. Natural occurrences are caused by arterial blockage, disease, and infection. Suppose that 20 patients will visit an emergency room with heart failure. Assume that causes of heart failure for the individuals are independent.

a. What is the probability that three individuals have conditions caused by outside factors?

b. What is the probability that three or more individuals have conditions caused by outside factors?

c. What are the mean and standard deviation of the number of individuals with conditions caused by outside factors?

3.5.11 A computer system uses passwords that are exactly six characters and each character is one of the 26 letters (a–z) or 10 integers (0–9). Suppose that 10,000 users of the system have unique passwords. A hacker randomly selects (with replacement) 100,000 passwords from the potential set, and a match to a user's password is called a *hit*.

a. What is the distribution of the number of hits?

b. What is the probability of no hits?

c. What are the mean and variance of the number of hits?

3.5.12 WP Samples of 20 parts from a metal punching process are selected every hour. Typically, 1% of the parts require rework. Let X denote the number of parts in the sample of 20 that require rework. A process problem is suspected if X exceeds its mean by more than 3 standard deviations.

a. If the percentage of parts that require rework remains at 1%, what is the probability that X exceeds its mean by more than 3 standard deviations?

b. If the rework percentage increases to 4%, what is the probability that X exceeds 1?

c. If the rework percentage increases to 4%, what is the probability that X exceeds 1 in at least one of the next five hours of samples?

3.5.13 WP Because all airline passengers do not show up for their reserved seat, an airline sells 125 tickets for a flight that holds only 120 passengers. The probability that a passenger does not show up is 0.10, and the passengers behave independently.

a. What is the probability that every passenger who shows up can take the flight?

b. What is the probability that the flight departs with empty seats?

3.5.14 WP This exercise illustrates that poor quality can affect schedules and costs. A manufacturing process has 100 customer orders to fill. Each order requires one component part that is purchased from a supplier. However, typically, 2% of the components are identified as defective, and the components can be assumed to be independent.

a. If the manufacturer stocks 100 components, what is the probability that the 100 orders can be filled without reordering components?

b. If the manufacturer stocks 102 components, what is the probability that the 100 orders can be filled without reordering components?

c. If the manufacturer stocks 105 components, what is the probability that the 100 orders can be filled without reordering components?

3.5.15 WP Consider the lengths of stay at a hospital's emergency department in Exercise 3.1.19. Assume that five persons independently arrive for service.

a. What is the probability that the length of stay of exactly one person is less than or equal to 4 hours?

b. What is the probability that exactly two people wait more than 4 hours?

c. What is the probability that at least one person waits more than 4 hours?

3.5.16 The probability that a visitor to a Web site provides contact data for additional information is 0.01. Assume that 1000 visitors to the site behave independently. Determine the following probabilities:

a. No visitor provides contact data.

b. Exactly 10 visitors provide contact data.

c. More than 3 visitors provide contact data.

3.5.17 Consider the time to recharge the flash in cell-phone cameras as in Example 3.1. Assume that the probability that a camera passes the test is 0.8 and the cameras perform independently. What is the smallest sample size needed so that the probability of at least one camera failing is at least 95%?

Exercises for Section 3.6

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

3.6.1 WP SS Suppose that the random variable X has a geometric distribution with $p = 0.5$. Determine the following probabilities:

a. $P(X = 1)$ **b.** $P(X = 4)$ **c.** $P(X = 8)$
d. $P(X \leq 2)$ **e.** $P(X > 2)$

3.6.2 WP Suppose that X is a negative binomial random variable with $p = 0.2$ and $r = 4$. Determine the following:

a. $E(X)$ **b.** $P(X = 20)$
c. $P(X = 19)$ **d.** $P(X = 21)$
e. The most likely value for X

3.6.3 WP Consider a sequence of independent Bernoulli trials with $p = 0.2$.

a. What is the expected number of trials to obtain the first success?
b. After the eighth success occurs, what is the expected number of trials to obtain the ninth success?

3.6.4 WP In a clinical study, volunteers are tested for a gene that has been found to increase the risk for a disease. The probability that a person carries the gene is 0.1. People are assumed to be independent with respect to the gene.

a. What is the probability that exactly four people need to be tested to detect two with the gene?
b. What is the probability that four or more people need to be tested to detect two with the gene?
c. What is the expected number of people to test to detect two with the gene?

3.6.5 WP Assume that each of your calls to a popular radio station has a probability of 0.02 of connecting, that is, of not obtaining a busy signal. Assume that your calls are independent.

a. What is the probability that your first call that connects is your 10th call?
b. What is the probability that it requires more than five calls for you to connect?
c. What is the mean number of calls needed to connect?

3.6.6 WP A player of a video game is confronted with a series of opponents and has an 80% probability of defeating each one. Success with any opponent is independent of previous encounters. Until defeated, the player continues to contest opponents.

a. What is the probability mass function of the number of opponents contested in a game?
b. What is the probability that a player defeats at least two opponents in a game?
c. What is the expected number of opponents contested in a game?
d. What is the probability that a player contests four or more opponents in a game?
e. What is the expected number of game plays until a player contests four or more opponents?

3.6.7 WP Heart failure is due to either natural occurrences (87%) or outside factors (13%). Outside factors are related to induced substances or foreign objects. Natural occurrences are caused by arterial blockage, disease, and infection. Assume that causes of heart failure for the individuals are independent.

a. What is the probability that the first patient with heart failure who enters the emergency room has the condition due to outside factors?
b. What is the probability that the third patient with heart failure who enters the emergency room is the first one due to outside factors?
c. What is the mean number of heart failure patients with the condition due to natural causes who enter the emergency room before the first patient with heart failure from outside factors?

3.6.8 Assume that 20 parts are checked each hour and that X denotes the number of parts in the sample of 20 that require rework. Parts are assumed to be independent with respect to rework.

a. If the percentage of parts that require rework remains at 1%, what is the probability that hour 10 is the first sample at which X exceeds 1?
b. If the rework percentage increases to 4%, what is the probability that hour 10 is the first sample at which X exceeds 1?
c. If the rework percentage increases to 4%, what is the expected number of hours until X exceeds 1?

3.6.9 SS VS A trading company uses eight computers to trade on the New York Stock Exchange (NYSE). The probability of a computer failing in a day is 0.005, and the computers fail independently. Computers are repaired in the evening, and each day is an independent trial.

a. What is the probability that all eight computers fail in a day?
b. What is the mean number of days until a specific computer fails?
c. What is the mean number of days until all eight computers fail on the same day?

3.6.10 WP In the process of meiosis, a single parent diploid cell goes through eight different phases. However, only 60% of the processes pass the first six phases and only 40% pass all eight. Assume that the results from each phase are independent.

a. If the probability of a successful pass of each one of the first six phases is constant, what is the probability of a successful pass of a single one of these phases?
b. If the probability of a successful pass of each one of the last two phases is constant, what is the probability of a successful pass of a single one of these phases?

3.6.11 A Web site randomly selects among 10 products to discount each day. The color printer of interest to you is discounted today.

a. What is the expected number of days until this product is again discounted?

b. What is the probability that this product is first discounted again exactly 10 days from now?

c. If the product is not discounted for the next five days, what is the probability that it is first discounted again 15 days from now?

d. What is the probability that this product is first discounted again within three or fewer days?

3.6.12 Consider the visits that result in leave without being seen (LWBS) at an emergency department in Example 2.6. Assume that people independently arrive for service at hospital l.

a. What is the probability that the fifth visit is the first one to LWBS?

b. What is the probability that either the fifth or sixth visit is the first one to LWBS?

c. What is the probability that the first visit to LWBS is among the first four visits?

d. What is the expected number of visits until the third LWBS occurs?

3.6.13 Consider the time to recharge the flash in cell-phone cameras as in Example 3.1. Assume that the probability that a camera passes the test is 0.8 and the cameras perform independently. Determine the following:

a. Probability that the second failure occurs on the 10th camera tested.

b. Probability that the second failure occurs in tests of four or fewer cameras.

c. Expected number of cameras tested to obtain the third failure.

3.6.14 Customers visit a Web site, and the probability of an order if a customer views five or fewer pages is 0.01. However, if a customer views more than five pages, the probability of an order is 0.1. The probability a customer views five or more pages is 0.25. The customers behave independently.

a. Is the number of customers who visit the site until an order is obtained a geometric random variable? Why or why not?

b. What is the probability that the first order is obtained from the 10th customer to visit the site?

3.6.15 An array of 30 LED bulbs is used in an automotive light. The probability that a bulb is defective is 0.001 and defective bulbs occur independently. Determine the following:

a. Probability that an automotive light has two or more defective bulbs.

b. Expected number of automotive lights to check to obtain one with two or more defective bulbs.

Exercises for Section 3.7

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

3.7.1 **WP** Suppose that X has a hypergeometric distribution with $N = 100$, $n = 4$, and $K = 20$. Determine the following:

a. $P(X = 1)$ **b.** $P(X = 6)$

c. $P(X = 4)$ **d.** Mean and variance of X

3.7.2 **WP** Suppose that X has a hypergeometric distribution with $N = 10$, $n = 3$, and $K = 4$. Sketch the probability mass function of X. Determine the cumulative distribution function for X.

3.7.3 **WP** **SS** A research study uses 800 men under the age of 55. Suppose that 30% carry a marker on the male chromosome that indicates an increased risk for high blood pressure.

a. If 10 men are selected randomly and tested for the marker, what is the probability that exactly 1 man has the marker?

b. If 10 men are selected randomly and tested for the marker, what is the probability that more than 1 has the marker?

3.7.4 **WP** Printed circuit cards are placed in a functional test after being populated with semiconductor chips. A lot contains 140 cards, and 20 are selected without replacement for functional testing.

a. If 20 cards are defective, what is the probability that at least 1 defective card is in the sample?

b. If 5 cards are defective, what is the probability that at least 1 defective card appears in the sample?

3.7.5 **WP** The analysis of results from a leaf transmutation experiment (turning a leaf into a petal) is summarized by the type of transformation completed:

		Total Textural Transformation	
		Yes	No
Total Color	Yes	243	26
Transformation	No	13	18

A naturalist randomly selects three leaves from this set without replacement. Determine the following probabilities.

a. Exactly one has undergone both types of transformations.

b. At least one has undergone both transformations.

c. Exactly one has undergone one but not both transformations.

d. At least one has undergone at least one transformation.

3.7.6 **WP** A state runs a lottery in which six numbers are randomly selected from 40 without replacement. A player chooses six numbers before the state's sample is selected.

a. What is the probability that the six numbers chosen by a player match all six numbers in the state's sample?

b. What is the probability that five of the six numbers chosen by a player appear in the state's sample?

c. What is the probability that four of the six numbers chosen by a player appear in the state's sample?

d. If a player enters one lottery each week, what is the expected number of weeks until a player matches all six numbers in the state's sample?

3.7.7 WP A slitter assembly contains 48 blades. Five blades are selected at random and evaluated each day for sharpness. If any dull blade is found, the assembly is replaced with a newly sharpened set of blades.

a. If 10 of the blades in an assembly are dull, what is the probability that the assembly is replaced the first day it is evaluated?

b. If 10 of the blades in an assembly are dull, what is the probability that the assembly is not replaced until the third day of evaluation? [*Hint*: Assume that the daily decisions are independent, and use the geometric distribution.]

c. Suppose that on the first day of evaluation, 2 of the blades are dull; on the second day of evaluation, 6 are dull; and on the third day of evaluation, 10 are dull. What is the probability that the assembly is not replaced until the third day of evaluation? [*Hint*: Assume that the daily decisions are independent. However, the probability of replacement changes every day.]

3.7.8 **a.** For Exercise 3.7.1, calculate $P(X = 1)$ and $P(X = 4)$, assuming that X has a binomial distribution, and compare these results to results derived from the hypergeometric distribution.

b. Use the binomial approximation to the hypergeometric distribution to approximate the probabilities in Exercise 3.7.4. What is the finite population correction in this exercise?

3.7.9 SS Suppose that a healthcare provider selects 20 patients randomly (without replacement) from among 500 to evaluate adherence to a medication schedule. Suppose that 10% of the 500 patients fail to adhere with the schedule. Determine the following:

a. Probability that exactly 10% of the patients in the sample fail to adhere.

b. Probability that fewer than 10% of the patients in the sample fail to adhere.

c. Probability that more than 10% of the patients in the sample fail to adhere.

d. Mean and variance of the number of patients in the sample who fail to adhere.

3.7.10 A utility company might offer electrical rates based on time-of-day consumption to decrease the peak demand in a day. Enough customers need to accept the plan for it to be successful. Suppose that among 50 major customers, 15 would accept the plan. The utility selects 10 major customers randomly (without replacement) to contact and promote the plan.

a. What is the probability that exactly two of the selected major customers accept the plan?

b. What is the probability that at least one of the selected major customers accepts the plan?

c. Instead of 15 customers, what is the minimum number of major customers that would need to accept the plan to meet the following objective? The probability that at least one selected major customer accepts the plan is greater than or equal to 0.95.

3.7.11 Suppose that lesions are present at 5 sites among 50 in a patient. A biopsy selects 8 sites randomly (without replacement).

a. What is the probability that lesions are present in at least one selected site?

b. What is the probability that lesions are present in two or more selected sites?

c. Instead of eight sites, what is the minimum number of sites that need to be selected to meet the following objective? The probability that at least one site has lesions present is greater than or equal to 0.9.

Exercises for Section 3.8

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

3.8.1 WP Suppose that X has a Poisson distribution. Determine the following probabilities when the mean of X is 4 and repeat for a mean of 0.4:

a. $P(X = 0)$ **b.** $P(X \leq 2)$
c. $P(X = 4)$ **d.** $P(X = 8)$

3.8.2 The number of telephone calls that arrive at a phone exchange is often modeled as a Poisson random variable. Assume that on the average there are 10 calls per hour. Determine the following probabilities:

a. exactly 5 calls in one hour
b. 3 or fewer calls in one hour
c. exactly 15 calls in two hours
d. exactly 5 calls in 30 minutes

3.8.3 WP Suppose that the number of customers who enter a store in an hour is a Poisson random variable, and suppose that $P(X = 0) = 0.05$. Determine the mean and variance of X.

3.8.4 Data from *www.centralhudsonlabs.com* determined the mean number of insect fragments in 225-gram chocolate bars was 14.4, but three brands had insect contamination more than twice the average. See the U.S. Food and Drug Administration–Center for Food Safety and Applied Nutrition for Defect Action Levels for food products. Assume that the number of fragments (contaminants) follows a Poisson distribution.

a. If you consume a 225-gram bar from a brand at the mean contamination level, what is the probability of no insect contaminants?

b. Suppose that you consume a bar that is one-fifth the size tested (45 grams) from a brand at the mean contamination level. What is the probability of no insect contaminants?

c. If you consume seven 28.35-gram (one-ounce) bars this week from a brand at the mean contamination level, what is

the probability that you consume one or more insect fragments in more than one bar?

d. Is the probability of contamination more than twice the mean of 14.4 unusual, or can it be considered typical variation? Explain.

3.8.5 WP SS VS Astronomers treat the number of stars in a given volume of space as a Poisson random variable. The density in the Milky Way Galaxy in the vicinity of our solar system is one star per 16 cubic light-years.

a. What is the probability of two or more stars in 16 cubic light-years?

b. How many cubic light-years of space must be studied so that the probability of one or more stars exceeds 0.95?

3.8.6 The number of cracks in a section of interstate highway that are significant enough to require repair is assumed to follow a Poisson distribution with a mean of two cracks per mile.

a. What is the probability that there are no cracks that require repair in 5 miles of highway?

b. What is the probability that at least one crack requires repair in 1/2 mile of highway?

c. If the number of cracks is related to the vehicle load on the highway and some sections of the highway have a heavy load of vehicles whereas other sections carry a light load, what do you think about the assumption of a Poisson distribution for the number of cracks that require repair?

3.8.7 WP In 1898, L. J. Bortkiewicz published a book titled *The Law of Small Numbers*. He used data collected over 20 years to show that the number of soldiers killed by horse kicks each year in each corps in the Prussian cavalry followed a Poisson distribution with a mean of 0.61.

a. What is the probability of more than one death in a corps in a year?

b. What is the probability of no deaths in a corps over five years?

3.8.8 The number of views of a page on a Web site follows a Poisson distribution with a mean of 1.5 per minute.

a. What is the probability of no views in a minute?

b. What is the probability of two or fewer views in 10 minutes?

c. Does the answer to the previous part depend on whether the 10-minute period is an uninterrupted interval? Explain.

d. Determine the length of a time interval such that the probability of no views in an interval of this length is 0.001.

3.8.9 WP The number of surface flaws in plastic panels used in the interior of automobiles has a Poisson distribution with a mean of 0.05 flaw per square foot of plastic panel. Assume that an automobile interior contains 10 square feet of plastic panel.

a. What is the probability that there are no surface flaws in an auto's interior?

b. If 10 cars are sold to a rental company, what is the probability that none of the 10 cars has any surface flaws?

c. If 10 cars are sold to a rental company, what is the probability that at most 1 car has any surface flaws?

3.8.10 *Inclusions* are defects in poured metal caused by contaminants. The number of (large) inclusions in cast iron follows a Poisson distribution with a mean of 2.5 per cubic millimeter. Determine the following:

a. Probability of at least one inclusion in a cubic millimeter.

b. Probability of at least five inclusions in 5.0 cubic millimeters.

c. Volume of material to inspect such that the probability of at least one inclusion is 0.99.

d. Instead of a mean of 2.5 per cubic millimeters, the mean inclusions per cubic millimeter such that the probability of at least one inclusion is 0.95.

3.8.11 Cabs pass your workplace according to a Poisson process with a mean of five cabs per hour. Suppose that you exit the workplace at 6:00 P.M. Determine the following:

a. Probability that you wait more than 10 minutes for a cab.

b. Probability that you wait less than 20 minutes for a cab.

c. Mean number of cabs per hour so that the probability that you wait more than 10 minutes is 0.1.

3.8.12 The article "An Association Between Fine Particles and Asthma Emergency Department Visits for Children in Seattle" [*Environmental Health Perspectives* June 1999, Vol. 107(6)] used Poisson models for the number of asthma emergency department (ED) visits per day. For the zip codes studied, the mean ED visits were 1.8 per day. Determine the following:

a. Probability of more than five visits in a day.

b. Probability of fewer than five visits in a week.

c. Number of days such that the probability of at least one visit is 0.99.

d. Instead of a mean of 1.8 per day, determine the mean visits per day such that the probability of more than five visits in a day is 0.1.

Supplemental Exercises for Chapter 3

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

3.S13 WP Let the random variable X be equally likely to assume any of the values 1/8, 1/4, or 3/8. Determine the mean and variance of X.

3.S14 WP Let X denote the number of bits received in error in a digital communication channel, and assume that X is a binomial random variable with $p = 0.001$. If 1000 bits are transmitted, determine the following:

a. $P(X = 1)$ **b.** $P(X \geq 1)$

c. $P(X \leq 2)$ **d.** mean and variance of X

3.S15 An automated egg carton loader has a 1% probability of cracking an egg, and a customer will complain if more than one

egg per dozen is cracked. Assume that each egg load is an independent event.

a. What is the distribution of cracked eggs per dozen? Include parameter values.

b. What is the probability that a carton of a dozen eggs results in a complaint?

c. What are the mean and standard deviation of the number of cracked eggs in a carton of a dozen eggs?

3.S16 A congested computer network has a 1% chance of losing a data packet that must be resent, and packet losses are independent events. An e-mail message requires 100 packets.

a. What is the distribution of the number of packets in an e-mail message that must be resent? Include the parameter values.

b. What is the probability that at least one packet is resent?

c. What is the probability that two or more packets are resent?

d. What are the mean and standard deviation of the number of packets that are resent?

e. If there are 10 messages and each contains 100 packets, what is the probability that at least one message requires that two or more packets be resent?

3.S17 WP SS A particularly long traffic light on your morning commute is green on 20% of the mornings. Assume that each morning represents an independent trial.

a. What is the probability that the first morning that the light is green is the fourth morning?

b. What is the probability that the light is not green for 10 consecutive mornings?

3.S18 The probability that an eagle kills a rabbit in a day of hunting is 10%. Assume that results are independent for each day.

a. What is the distribution of the number of days until a successful hunt?

b. What is the probability that the first successful hunt occurs on day five?

c. What is the expected number of days until a successful hunt?

d. If the eagle can survive up to 10 days without food (it requires a successful hunt on the 10th day), what is the probability that the eagle is still alive 10 days from now?

3.S19 WP An electronic scale in an automated filling operation stops the manufacturing line after three underweight packages are detected. Suppose that the probability of an underweight package is 0.001 and each fill is independent.

a. What is the mean number of fills before the line is stopped?

b. What is the standard deviation of the number of fills before the line is stopped?

3.S20 WP A shipment of chemicals arrives in 15 totes. Three of the totes are selected at random without replacement for an inspection of purity. If two of the totes do not conform to purity requirements, what is the probability that at least one of the nonconforming totes is selected in the sample?

3.S21 Traffic flow is traditionally modeled as a Poisson distribution. A traffic engineer monitors the traffic flowing through an intersection with an average of six cars per minute. To set the timing of a traffic signal, the following probabilities are used.

a. What is the probability that no cars pass through the intersection within 30 seconds?

b. What is the probability that three or more cars pass through the intersection within 30 seconds?

c. Calculate the minimum number of cars through the intersection so that the probability of this number or fewer cars in 30 seconds is at least 90%.

d. If the variance of the number of cars through the intersection per minute is 20, is the Poisson distribution appropriate? Explain.

3.S22 WP The number of messages that arrive at a Web site is a Poisson random variable with a mean of five messages per hour.

a. What is the probability that five messages are received in 1.0 hour?

b. What is the probability that 10 messages are received in 1.5 hours?

c. What is the probability that fewer than two messages are received in 0.5 hour?

d. Determine the length of an interval of time such that the probability that no messages arrive during this interval is 0.90.

3.S23 WP The probability that your call to a service line is answered in less than 30 seconds is 0.75. Assume that your calls are independent.

a. If you call 10 times, what is the probability that exactly nine of your calls are answered within 30 seconds?

b. If you call 20 times, what is the probability that at least 16 calls are answered in less than 30 seconds?

c. If you call 20 times, what is the mean number of calls that are answered in less than 30 seconds?

d. What is the probability that you must call four times to obtain the first answer in less than 30 seconds?

e. What is the mean number of calls until you are answered in less than 30 seconds?

3.S24 WP The number of errors in a textbook follows a Poisson distribution with a mean of 0.01 error per page. What is the probability that there are three or fewer errors in 100 pages?

3.S25 The probability that an individual recovers from an illness in a one-week time period without treatment is 0.1. Suppose that 20 independent individuals suffering from this illness are treated with a drug and 4 recover in a one-week time period. If the drug has no effect, what is the probability that 4 or more people recover in a one-week time period?

3.S26 Patient response to a generic drug to control pain is scored on a 5-point scale where a 5 indicates complete relief. Historically, the distribution of scores is

1	2	3	4	5
0.05	0.1	0.2	0.25	0.4

Two patients, assumed to be independent, are each scored.

a. What is the probability mass function of the total score?

b. What is the probability mass function of the average score?

3.S27 WP In a manufacturing process that laminates several ceramic layers, 1% of the assemblies are defective. Assume that the assemblies are independent.

a. What is the mean number of assemblies that need to be checked to obtain five defective assemblies?

b. What is the standard deviation of the number of assemblies that need to be checked to obtain five defective assemblies?

c. Determine the minimum number of assemblies that need to be checked so that the probability that at least one defective assembly is obtained exceeds 0.95.

3.S28 WP A manufacturer of a consumer electronics product expects 2% of units to fail during the warranty period. A sample of 500 independent units is tracked for warranty performance.

a. What is the probability that none fails during the warranty period?

b. What is the expected number of failures during the warranty period?

c. What is the probability that more than two units fail during the warranty period?

3.S29 Determine the probability mass function for the random variable with the following cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < 2 \\ 0.2 & 2 \le x < 5.7 \\ 0.5 & 5.7 \le x < 6.5 \\ 0.8 & 6.5 \le x < 8.5 \\ 1 & 8.5 \le x \end{cases}$$

3.S30 WP VS The random variable X has the following probability distribution:

x	2	3	5	8
Probability	0.2	0.4	0.3	0.1

Determine the following:

a. $P(X \le 3)$ **b.** $P(X > 2.5)$

c. $P(2.7 < X < 5.1)$ **d.** $E(X)$ **e.** $V(X)$

3.S31 SS Assume that the number of errors along a magnetic recording surface is a Poisson random variable with a mean of one error every 10^5 bits. A sector of data consists of 4096 eight-bit bytes.

a. What is the probability of more than one error in a sector?

b. What is the mean number of sectors until an error occurs?

3.S32 WP Each main bearing cap in an engine contains 4 bolts. The bolts are selected at random without replacement from a parts bin that contains 30 bolts from one supplier and 70 bolts from another.

a. What is the probability that a main bearing cap contains all bolts from the same supplier?

b. What is the probability that exactly 3 bolts are from the same supplier?

3.S33 From 500 customers, a major appliance manufacturer randomly selects a sample without replacement. The company estimates that 25% of the customers will reply to the survey. If this estimate is correct, what is the probability mass function of the number of customers that will reply?

a. Assume that the company samples 5 customers.

b. Assume that the company samples 10 customers.

3.S34 WP SS An installation technician for a specialized communication system is dispatched to a city only when three or more orders have been placed. Suppose that orders follow a Poisson distribution with a mean of 0.25 per week for a city with a population of 100,000, and suppose that your city contains a population of 800,000.

a. What is the probability that a technician is required after a one-week period?

b. If you are the first one in the city to place an order, what is the probability that you have to wait more than two weeks from the time you place your order until a technician is dispatched?

3.S35 Saguaro cacti are large cacti indigenous to the southwestern United States and Mexico. Assume that the number of saguaro cacti in a region follows a Poisson distribution with a mean of 280 per square kilometer. Determine the following:

a. Mean number of cacti per 10,000 square meters.

b. Probability of no cacti in 10,000 square meters.

c. Area of a region such that the probability of at least two cacti in the region is 0.9.

3.S36 WP It is suspected that some of the totes containing chemicals purchased from a supplier exceed the moisture content target. Assume that the totes are independent with respect to moisture content. Determine the proportion of totes from the supplier that must exceed the moisture content target so that the probability is 0.90 that at least 1 tote in a sample of 30 fails the test.

Chapter 4 Exercises

Exercises for Section 4.1

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

4.1.1 WP Suppose that $f(x) = e^{-x}$ for $0 < x$. Determine the following:

a. $P(1 < X)$ **b.** $P(1 < X < 2.5)$ **c.** $P(X = 3)$

d. $P(X < 4)$ **e.** $P(3 \le X)$

f. x such that $P(x < X) = 0.10$

g. x such that $P(X \le x) = 0.10$

4.1.2 WP The diameter of a particle of contamination (in micrometers) is modeled with the probability density function $f(x) = 2/x^3$ for $x > 1$. Determine the following:

a. $P(X < 2)$ **b.** $P(X > 5)$
c. $P(4 < X < 8)$ **d.** $P(X < 4 \text{ or } X > 8)$
e. x such that $P(X < x) = 0.95$

4.1.3 SS Suppose that $f(x) = 0.5 \cos x$ for $-\pi/2 < x < \pi/2$. Determine the following:

a. $P(X < 0)$ **b.** $P(X < -\pi/4)$
c. $P(-\pi/4 < X < \pi/4)$ **d.** $P(X > -\pi/4)$
e. x such that $P(X < x) = 0.95$

4.1.4 WP GO Tutorial Suppose that $f(x) = x/8$ for $3 < x < 5$. Determine the following probabilities:

a. $P(X < 4)$ **b.** $P(X > 3.5)$
c. $P(4 < X < 5)$ **d.** $P(X < 4.5)$
e. $P(X < 3.5 \text{ or } X > 4.5)$

4.1.5 WP Suppose that $f(x) = 1.5x^2$ for $-1 < x < 1$. Determine the following:

a. $P(0 < X)$ **b.** $P(0.5 < X)$
c. $P(-0.5 \le X \le 0.5)$ **d.** $P(X < -2)$
e. $P(X < 0 \text{ or } X > -0.5)$
f. x such that $P(x < X) = 0.05$.

4.1.6 WP The probability density function of the net weight in pounds of a packaged chemical herbicide is $f(x) = 2.0$ for $49.75 < x < 50.25$ pounds.

a. Determine the probability that a package weighs more than 50 pounds.
b. How much chemical is contained in 90% of all packages?

4.1.7 WP SS The probability density function of the length of a metal rod is $f(x) = 2$ for $2.3 < x < 2.8$ meters.

a. If the specifications for this process are from 2.25 to 2.75 meters, what proportion of rods fail to meet the specifications?
b. Assume that the probability density function is $f(x) = 2$ for an interval of length 0.5 meters. Over what value should the density be centered to achieve the greatest proportion of rods within specifications?

4.1.8 An article in *Electric Power Systems Research* ["Modeling Real-Time Balancing Power Demands in Wind Power Systems Using Stochastic Differential Equations" (2010, Vol. 80(8), pp. 966–974)] considered a new probabilistic model to balance power demand with large amounts of wind power. In this model, the power loss from shutdowns is assumed to have a triangular distribution with probability density function

$$f(x) = \begin{cases} -5.56 \times 10^{-4} + 5.56 \times 10^{-6}x, & x \in [100,\ 500] \\ 4.44 \times 10^{-3} - 4.44 \times 10^{-6}x, & x \in [500, 1000] \\ 0, & \text{otherwise} \end{cases}$$

Determine the following:

a. $P(X < 90)$ **b.** $P(100 < X \le 200)$
c. $P(X > 800)$ **d.** Value exceeded with probability 0.1.

4.1.9 The waiting time for service at a hospital emergency department (in hours) follows a distribution with probability density function $f(x) = 0.5 \exp(-0.5x)$ for $0 < x$. Determine the following:

a. $P(X < 0.5)$ **b.** $P(X > 2)$
c. Value x (in hours) exceeded with probability 0.05.

4.1.10 The distribution of X is approximated with a triangular probability density function $f(x) = 0.0025x - 0.075$ for $30 < x < 50$ and $f(x) = -0.0025x + 0.175$ for $50 < x < 70$. Determine the following:

a. $P(X \le 40)$ **b.** $P(40 < X \le 60)$
c. Value x exceeded with probability 0.99.

Exercises for Section 4.2

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

4.2.1 WP Suppose that the cumulative distribution function of the random variable X is

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.25x & 0 \le x < 5 \\ 1 & 5 \le x \end{cases}$$

Determine the following:

a. $P(X < 2.8)$ **b.** $P(X > 1.5)$
c. $P(X < -2)$ **d.** $P(X > 6)$

4.2.2 WP Determine the cumulative distribution function for the distribution in Exercise 4.1.1.

4.2.3 Determine the cumulative distribution function for the distribution in Exercise 4.1.3.

4.2.4 Determine the cumulative distribution function for the distribution in Exercise 4.1.2.

4.2.5 SS Determine the cumulative distribution function for the distribution in Exercise 4.1.4.

4.2.6 WP The probability density function of the time you arrive at a terminal (in minutes after 8:00 A.M.) is $f(x) = 0.1 \exp(-0.1x)$ for $0 < x$. Determine the probability that

a. You arrive by 9:00 A.M.
b. You arrive between 8:15 A.M. and 8:30 A.M.
c. You arrive before 8:40 A.M. on two or more days of five days. Assume that your arrival times on different days are independent.
d. Determine the cumulative distribution function and use the cumulative distribution function to determine the probability that you arrive between 8:15 A.M. and 8:30 A.M.

4.2.7 Determine the cumulative distribution function for the distribution in Exercise 4.1.7. Use the cumulative distribution function to determine the probability that a length exceeds 2.7 meters.

Determine the probability density function for each of the following cumulative distribution functions.

4.2.8 WP $F(x) = 1 - e^{-2x} \quad x > 0$

4.2.9 WP SS

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.2x & 0 \le x < 4 \\ 0.04x + 0.64 & 4 \le x < 9 \\ 1 & 9 \le x \end{cases}$$

4.2.10 Determine the cumulative distribution function for the random variable in Exercise 4.1.10. Use the cumulative distribution function to determine the probability that the random variable is less than 55.

4.2.11 Determine the cumulative distribution function for the random variable in Exercise 4.1.9. Use the cumulative distribution function to determine the probability that the waiting time is less than 1 hour.

4.2.12 Determine the cumulative distribution function for the random variable in Exercise 4.1.8. Use the cumulative distribution function to determine the probability that $400 < X < 500$.

Exercises for Section 4.3

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

4.3.1 WP SS VS Suppose that $f(x) = 1.5x^2$ for $-1 < x < 1$. Determine the mean and variance of X.

4.3.2 WP Suppose that $f(x) = 0.125x$ for $0 < x < 4$. Determine the mean and variance of X.

4.3.3 Determine the mean and variance of the random variable in Exercise 4.1.1.

4.3.4 Determine the mean and variance of the random variable in Exercise 4.1.8.

4.3.5 Determine the mean and variance of the random variable in Exercise 4.1.10.

4.3.6 WP Suppose that the probability density function of the length of computer cables is $f(x) = 0.1$ from 1200 to 1210 millimeters.

a. Determine the mean and standard deviation of the cable length.

b. If the length specifications are $1195 < x < 1205$ millimeters, what proportion of cables is within specifications?

4.3.7 WP SS The thickness of a conductive coating in micrometers has a density function of $600x^{-2}$ for $100\ \mu m < x < 120\ \mu m$.

a. Determine the mean and variance of the coating thickness.

b. If the coating costs \$0.50 per micrometer of thickness on each part, what is the average cost of the coating per part?

Exercises for Section 4.4

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

4.4.1 WP Suppose X has a continuous uniform distribution over the interval $[-1, 1]$. Determine the following:

a. Mean, variance, and standard deviation of X

b. Value for x such that $P(-x < X < x) = 0.90$

c. Cumulative distribution function

4.4.2 WP VS The thickness of a flange on an aircraft component is uniformly distributed between 0.95 and 1.05 millimeters. Determine the following:

a. Cumulative distribution function of flange thickness

b. Proportion of flanges that exceeds 1.02 millimeters

c. Thickness exceeded by 90% of the flanges

d. Mean and variance of flange thickness

4.4.3 WP The thickness of photoresist applied to wafers in semiconductor manufacturing at a particular location on the wafer is uniformly distributed between 0.2050 and 0.2150 micrometers. Determine the following:

a. Cumulative distribution function of photoresist thickness

b. Proportion of wafers that exceeds 0.2125 micrometers in photoresist thickness

c. Thickness exceeded by 10% of the wafers

d. Mean and variance of photoresist thickness

4.4.4 An adult can lose or gain two pounds of water in the course of a day. Assume that the changes in water weight are uniformly distributed between minus two and plus two pounds in a day. What is the standard deviation of a person's weight over a day?

4.4.5 WP SS A show is scheduled to start at 9:00 A.M., 9:30 A.M., and 10:00 A.M. Once the show starts, the gate will be closed. A visitor will arrive at the gate at a time uniformly distributed between 8:30 A.M. and 10:00 A.M. Determine the following:

a. Cumulative distribution function of the time (in minutes) between arrival and 8:30 A.M.

b. Mean and variance of the distribution in the previous part

c. Probability that a visitor waits less than 10 minutes for a show

d. Probability that a visitor waits more than 20 minutes for a show

4.4.6 The volume of a shampoo filled into a container is uniformly distributed between 374 and 380 milliliters.

a. What are the mean and standard deviation of the volume of shampoo?

b. What is the probability that the container is filled with less than the advertised target of 375 milliliters?

c. What is the volume of shampoo that is exceeded by 95% of the containers?

d. Every milliliter of shampoo costs the producer \$0.002. Any shampoo more than 375 milliliters in the container is an extra cost to the producer. What is the mean extra cost?

4.4.7 An e-mail message will arrive at a time uniformly distributed between 9:00 A.M. and 11:00 A.M. You check e-mail at 9:15 A.M. and every 30 minutes afterward.

a. What is the standard deviation of arrival time (in minutes)?

b. What is the probability that the message arrives less than 10 minutes before you view it?

c. What is the probability that the message arrives more than 15 minutes before you view it?

4.4.8 An electron emitter produces electron beams with changing kinetic energy that is uniformly distributed between 3 and 7 joules. Suppose that it is possible to adjust the upper limit of the kinetic energy (currently set to 7 joules).

a. What is the mean kinetic energy?

b. What is the variance of the kinetic energy?

c. What is the probability that an electron beam has a kinetic energy of exactly 3.2 joules?

d. What should be the upper limit so that the mean kinetic energy increases to 8 joules?

e. What should be the upper limit so that the variance of kinetic energy decreases to 0.75 joules?

Exercises for Section 4.5

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

4.5.1 **WP** Use Appendix Table III to determine the following probabilities for the standard normal random variable Z:

a. $P(Z < 1.32)$ **b.** $P(Z < 3.0)$
c. $P(Z > 1.45)$ **d.** $P(Z > -2.15)$
e. $P(-2.34 < Z < 1.76)$

4.5.2 **WP** Assume that Z has a standard normal distribution. Use Appendix Table III to determine the value for z that solves each of the following:

a. $P(-z < Z < z) = 0.95$ **b.** $P(-z < Z < z) = 0.99$
c. $P(-z < Z < z) = 0.68$ **d.** $P(-z < Z < z) = 0.9973$

4.5.3 **WP** **SS** Assume that X is normally distributed with a mean of 10 and a standard deviation of 2. Determine the following:

a. $P(Z < 13)$ **b.** $P(Z > 9)$
c. $P(6 < X < 14)$ **d.** $P(2 < X < 4)$
e. $P(-2 < X < 8)$

4.5.4 **WP** Assume that X is normally distributed with a mean of 10 and a standard deviation of 2. Determine the value for x that solves each of the following:

a. $P(X > x) = 0.5$ **b.** $P(X > x) = 0.95$
c. $P(x < X < 10) = 0$ **d.** $P(-x < X - 10 < x) = 0.95$
e. $P(-x < X - 10 < x) = 0.99$

4.5.5 An article in *Knee Surgery Sports Traumatology Arthroscopy* ["Effect of Provider Volume on Resource Utilization for Surgical Procedures" (2005, Vol. 13, pp. 273–279)] showed a mean time of 129 minutes and a standard deviation of 14 minutes for anterior cruciate ligament (ACL) reconstruction surgery at high-volume hospitals (with more than 300 such surgeries per year).

a. What is the probability that your ACL surgery at a high-volume hospital requires a time more than 2 standard deviations above the mean?

b. What is the probability that your ACL surgery at a high-volume hospital is completed in less than 100 minutes?

c. The probability of a completed ACL surgery at a high-volume hospital is equal to 95% at what time?

d. If your surgery requires 199 minutes, what do you conclude about the volume of such surgeries at your hospital? Explain.

4.5.6 **WP** The time until recharge for a battery in a laptop computer under common conditions is normally distributed with a mean of 260 minutes and a standard deviation of 50 minutes.

a. What is the probability that a battery lasts more than four hours?

b. What are the quartiles (the 25% and 75% values) of battery life?

c. What value of life in minutes is exceeded with 95% probability?

4.5.7 **WP** **SS** **VS** A driver's reaction time to visual stimulus is normally distributed with a mean of 0.4 seconds and a standard deviation of 0.05 seconds.

a. What is the probability that a reaction requires more than 0.5 seconds?

b. What is the probability that a reaction requires between 0.4 and 0.5 seconds?

c. What reaction time is exceeded 90% of the time?

4.5.8 Cholesterol is a fatty substance that is an important part of the outer lining (membrane) of cells in the body of animals. Its normal range for an adult is 120–240 mg/dl. The Food and Nutrition Institute of the Philippines found that the total cholesterol level for Filipino adults has a mean of 159.2 mg/dl and

84.1% of adults have a cholesterol level less than 200 mg/dl (*www.fnri.dost.gov.ph/*). Suppose that the total cholesterol level is normally distributed.

a. Determine the standard deviation of this distribution.

b. What are the quartiles (the 25% and 75% percentiles) of this distribution?

c. What is the value of the cholesterol level that exceeds 90% of the population?

d. An adult is at moderate risk if cholesterol level is more than 1 but less than 2 standard deviations above the mean. What percentage of the population is at moderate risk according to this criterion?

e. An adult whose cholesterol level is more than 2 standard deviations above the mean is thought to be at high risk. What percentage of the population is at high risk?

f. An adult whose cholesterol level is less than 1 standard deviation below the mean is thought to be at low risk. What percentage of the population is at low risk?

4.5.9 In an accelerator center, an experiment needs a 1.41-cm-thick aluminum cylinder (*http://puhep1.princeton.edu/mumu/target/Solenoid_Coil.pdf*). Suppose that the thickness of a cylinder has a normal distribution with a mean of 1.41 cm and a standard deviation of 0.01 cm.

a. What is the probability that a thickness is greater than 1.42 cm?

b. What thickness is exceeded by 95% of the samples?

c. If the specifications require that the thickness is between 1.39 cm and 1.43 cm, what proportion of the samples meets specifications?

4.5.10 In 2002, the average height of a woman aged 20–74 years was 64 inches, with an increase of approximately 1 inch from 1960 (*http://usgovinfo.about.com/od/healthcare*). Suppose the height of a woman is normally distributed with a standard deviation of 2 inches.

a. What is the probability that a randomly selected woman in this population is between 58 inches and 70 inches?

b. What are the quartiles of this distribution?

c. Determine the height that is symmetric about the mean that includes 90% of this population.

d. What is the probability that five women selected at random from this population all exceed 68 inches?

4.5.11 The life of a semiconductor laser at a constant power is normally distributed with a mean of 7000 hours and a standard deviation of 600 hours.

a. What is the probability that a laser fails before 5000 hours?

b. What is the life in hours that 95% of the lasers exceed?

c. If three lasers are used in a product and they are assumed to fail independently, what is the probability that all three are still operating after 7000 hours?

4.5.12 **WP GO Tutorial** The demand for water use in Phoenix in 2003 hit a high of about 442 million gallons per day on June 27 (*http://phoenix.gov/WATER/wtrfacts.html*). Water use in the summer is normally distributed with a mean of 310 million gallons per day and a standard deviation of 45 million gallons per day. City reservoirs have a combined storage capacity of nearly 350 million gallons.

a. What is the probability that a day requires more water than is stored in city reservoirs?

b. What reservoir capacity is needed so that the probability that it is exceeded is 1%?

c. What amount of water use is exceeded with 95% probability?

d. Water is provided to approximately 1.4 million people. What is the mean daily consumption per person at which the probability that the demand exceeds the current reservoir capacity is 1%? Assume that the standard deviation of demand remains the same.

4.5.13 **SS** The weight of a running shoe is normally distributed with a mean of 12 ounces and a standard deviation of 0.5 ounce.

a. What is the probability that a shoe weighs more than 13 ounces?

b. What must the standard deviation of weight be in order for the company to state that 99.9% of its shoes weighs less than 13 ounces?

c. If the standard deviation remains at 0.5 ounce, what must the mean weight be for the company to state that 99.9% of its shoes weighs less than 13 ounces?

4.5.14 The diameter of the dot produced by a printer is normally distributed with a mean diameter of 0.002 inch and a standard deviation of 0.0004 inch.

a. What is the probability that the diameter of a dot exceeds 0.0026?

b. What is the probability that a diameter is between 0.0014 and 0.0026?

c. What standard deviation of diameters is needed so that the probability in part (b) is 0.995?

4.5.15 Assume that a random variable is normally distributed with a mean of 24 and a standard deviation of 2. Consider an interval of length one unit that starts at the value a so that the interval is $[a, a + 1]$. For what value of a is the probability of the interval greatest? Does the standard deviation affect that choice of interval?

4.5.16 The length of stay at a specific emergency department in Phoenix, Arizona, in 2009 had a mean of 4.6 hours with a standard deviation of 2.9. Assume that the length of stay is normally distributed.

a. What is the probability of a length of stay greater than 10 hours?

b. What length of stay is exceeded by 25% of the visits?

c. From the normally distributed model, what is the probability of a length of stay less than 0 hours? Comment on the normally distributed assumption in this example.

4.5.17 An article in *Atmospheric Chemistry and Physics* ["Relationship Between Particulate Matter and Childhood Asthma—Basis of a Future Warning System for Central Phoenix" (2012, Vol. 12, pp. 2479–2490)] reported the use of PM10 (particulate matter <10 μm diameter) air quality data measured hourly from sensors in Phoenix, Arizona. The 24-hour (daily) mean PM10 for a centrally located sensor was 50.9 $\mu g/m^3$ with a standard deviation of 25.0. Assume that the daily mean of PM10 is normally distributed.

a. What is the probability of a daily mean of PM10 greater than 100 $\mu g/m^3$?

b. What is the probability of a daily mean of PM10 less than 25 $\mu g/m^3$?

c. What daily mean of PM10 value is exceeded with probability 5%?

4.5.18 An article in *Microelectronics Reliability* ["Advanced Electronic Prognostics through System Telemetry and Pattern Recognition Methods" (2007, Vol. 47(12), pp. 1865–1873)] presented an example of electronic prognosis. The objective was to detect faults to decrease the system downtime and the number of unplanned repairs in high-reliability systems. Previous measurements of the power supply indicated that the signal is normally distributed with a mean of 1.5 V and a standard deviation of 0.02 V.

a. Suppose that lower and upper limits of the predetermined specifications are 1.45 V and 1.55 V, respectively. What is the probability that a signal is within these specifications?

b. What is the signal value that is exceeded with 95% probability?

c. What is the probability that a signal value exceeds the mean by two or more standard deviations?

4.5.19 A signal in a communication channel is detected when the voltage is higher than 1.5 volts in absolute value. Assume that the voltage is normally distributed with a mean of 0. What is the standard deviation of voltage such that the probability of a false signal is 0.005?

4.5.20 An article in the *Journal of Cardiovascular Magnetic Resonance* ["Right Ventricular Ejection Fraction Is Better Reflected by Transverse Rather Than Longitudinal Wall Motion in Pulmonary Hypertension" (2010, Vol. 12(35)] discussed a study of the regional right-ventricle transverse wall motion in patients with pulmonary hypertension (PH). The right-ventricle ejection fraction (EF) was approximately normally distributed with a mean and a standard deviation of 36 and 12, respectively, for PH subjects, and with mean and standard deviation of 56 and 8, respectively, for control subjects.

a. What is the EF for PH subjects exceeded with 5% probability?

b. What is the probability that the EF of a control subject is less than the value in part (a)?

c. Comment on how well the control and PH subjects can be distinguished by EF measurements.

Exercises for Section 4.6

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

4.6.1 **WP** **SS** Suppose that X is a binomial random variable with $n = 200$ and $p = 0.4$. Approximate the following probabilities:

a. $P(X \leq 70)$ **b.** $P(70 < X < 90)$ **c.** $P(X = 80)$

4.6.2 **WP** Suppose that X is a Poisson random variable with $\lambda = 6$.

a. Compute the exact probability that X is less than four.

b. Approximate the probability that X is less than four and compare to the result in part (a).

c. Approximate the probability that $8 < X < 12$.

4.6.3 **WP** There were 49.7 million people with some type of long-lasting condition or disability living in the United States in 2000. This represented 19.3 percent of the majority of civilians aged five and over (*http://factfinder.census.gov*). A sample of 1000 persons is selected at random.

a. Approximate the probability that more than 200 persons in the sample have a disability.

b. Approximate the probability that between 180 and 300 people in the sample have a disability.

4.6.4 **WP** The manufacturing of semiconductor chips produces 2% defective chips. Assume that the chips are independent and that a lot contains 1000 chips. Approximate the following probabilities:

a. More than 25 chips are defective.

b. Between 20 and 30 chips are defective.

4.6.5 **VS** Hits to a high-volume Web site are assumed to follow a Poisson distribution with a mean of 10,000 per day. Approximate each of the following:

a. Probability of more than 20,000 hits in a day.

b. Probability of less than 9900 hits in a day.

c. Value such that the probability that the number of hits in a day exceeds the value is 0.01.

d. Expected number of days in a year (365 days) that exceed 10,200 hits.

e. Probability that over a year (365 days), each of the more than 15 days has more than 10,200 hits.

4.6.6 Phoenix water is provided to approximately 1.4 million people who are served through more than 362,000 accounts (*http://phoenix.gov/WATER/wtrfacts.html*). All accounts are metered and billed monthly. The probability that an account has an error in a month is 0.001, and accounts can be assumed to be independent.

a. What are the mean and standard deviation of the number of account errors each month?

b. Approximate the probability of fewer than 350 errors in a month.

c. Approximate a value so that the probability that the number of errors exceeds this value is 0.05.

d. Approximate the probability of more than 400 errors per month in the next two months. Assume that results between months are independent.

4.6.7 **SS** An article in *Atmospheric Chemistry and Physics* ["Relationship Between Particulate Matter and Childhood Asthma—Basis of a Future Warning System for Central Phoenix" (2012, Vol. 12, pp. 2479–2490)] linked air quality to childhood asthma incidents. The study region in central Phoenix, Arizona,

recorded 10,500 asthma incidents in children in a 21-month period. Assume that the number of asthma incidents follows a Poisson distribution.

a. Approximate the probability of more than 550 asthma incidents in a month.

b. Approximate the probability of 450 to 550 asthma incidents in a month.

c. Approximate the number of asthma incidents exceeded with probability 5%.

d. If the number of asthma incidents was greater during the winter than the summer, what would this imply about the Poisson distribution assumption?

4.6.8 An acticle in *Biometrics* ["Integrative Analysis of Transcriptomic and Proteomic Data of *Desulfovibrio Vulgaris:* A Nonlinear Model to Predict Abundance of Undetected Proteins" (2009, Vol. 25(15), pp. 1905–1914)] reported that protein abundance from an operon (a set of biologically related genes) was less dispersed than from randomly selected genes. In the research, 1000 sets of genes were randomly constructed, and of these sets, 75% were more disperse than a specific opteron. If the probability that a random set is more disperse than this opteron is truly 0.5, approximate the probability that 750 or more random sets exceed the opteron. From this result, what do you conclude about the dispersion in the opteron versus random genes?

4.6.9 A set of 200 independent patients take antiacid medication at the start of symptoms, and 80 experience moderate to substantial relief within 90 minutes. Historically, 30% of patients experience relief within 90 minutes with no medication. If the medication has no effect, approximate the probability that 80 or more patients experience relief of symptoms. What can you conclude about the effectiveness of this medication?

4.6.10 Cabs pass your workplace according to a Poisson process with a mean of five cabs per hour.

a. Determine the mean and standard deviation of the number of cabs per 10-hour day.

b. Approximate the probability that more than 65 cabs pass within a 10-hour day.

c. Approximate the probability that between 50 and 65 cabs pass in a 10-hour day.

d. Determine the mean hourly rate so that the probability is approximately 0.95 that 100 or more cabs pass in a 10-hour day.

4.6.11 The number of (large) inclusions in cast iron follows a Poisson distribution with a mean of 2.5 per cubic millimeter. Approximate the following probabilities:

a. Determine the mean and standard deviation of the number of inclusions in a cubic centimeter (cc).

b. Approximate the probability that fewer than 2600 inclusions occur in a cc.

c. Approximate the probability that more than 2400 inclusions occur in a cc.

d. Determine the mean number of inclusions per cubic millimeter such that the probability is approximately 0.9 that 500 or fewer inclusions occur in a cc.

Exercises for Section 4.7

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

4.7.1 **WP** **SS** Suppose that X has an exponential distribution with mean equal to 10. Determine the following:

a. $P(X > 10)$ **b.** $P(X > 20)$ **c.** $P(X < 30)$

d. Find the value of x such that $P(X < x) = 0.95$.

4.7.2 **WP** Suppose that X has an exponential distribution with a mean of 10. Determine the following:

a. $P(X < 5)$ **b.** $P(X < 15 \mid X > 10)$

c. Compare the results in parts (a) and (b) and comment on the role of the lack of memory property.

4.7.3 **WP** Suppose that the log-ons to a computer network follow a Poisson process with an average of three counts per minute.

a. What is the mean time between counts?

b. What is the standard deviation of the time between counts?

c. Determine x such that the probability that at least one count occurs before time x minutes is 0.95.

4.7.4 **WP** The life of automobile voltage regulators has an exponential distribution with a mean life of 6 years. You purchase a 6-year-old automobile with a working voltage regulator and plan to own it for 6 years.

a. What is the probability that the voltage regulator fails during your ownership?

b. If your regulator fails after you own the automobile 3 years and it is replaced, what is the mean time until the next failure?

4.7.5 **WP** **SS** **VS** The time between calls to a plumbing supply business is exponentially distributed with a mean time between calls of 15 minutes.

a. What is the probability that there are no calls within a 30-minute interval?

b. What is the probability that at least one call arrives within a 10-minute interval?

c. What is the probability that the first call arrives within 5 to 10 minutes after opening?

d. Determine the length of an interval of time such that the probability of at least one call in the interval is 0.90.

4.7.6 The number of stork sightings on a route in South Carolina follows a Poisson process with a mean of 2.3 per year.

a. What is the mean time between sightings?

b. What is the probability that there are no sightings within three months (0.25 years)?

c. What is the probability that the time until the first sighting exceeds six months?

d. What is the probability of no sighting within 3 years?

4.7.7 **WP** **VS** The time between arrivals of taxis at a busy intersection is exponentially distributed with a mean of 10 minutes.

a. What is the probability that you wait longer than 1 hour for a taxi?

b. Suppose that you have already been waiting for 1 hour for a taxi. What is the probability that one arrives within the next 10 minutes?

c. Determine x such that the probability that you wait more than x minutes is 0.10.

d. Determine x such that the probability that you wait less than x minutes is 0.90.

e. Determine x such that the probability that you wait less than x minutes is 0.50.

4.7.8 WP The distance between major cracks in a highway follows an exponential distribution with a mean of 5 miles.

a. What is the probability that there are no major cracks in a 10-mile stretch of the highway?

b. What is the probability that there are two major cracks in a 10-mile stretch of the highway?

c. What is the standard deviation of the distance between major cracks?

d. What is the probability that the first major crack occurs between 12 and 15 miles of the start of inspection?

e. What is the probability that there are no major cracks in two separate 5-mile stretches of the highway?

f. Given that there are no cracks in the first 5 miles inspected, what is the probability that there are no major cracks in the next 10 miles inspected?

4.7.9 According to results from the analysis of chocolate bars in Chapter 3, the mean number of insect fragments was 14.4 in 225 grams. Assume that the number of fragments follows a Poisson distribution.

a. What is the mean number of grams of chocolate until a fragment is detected?

b. What is the probability that there are no fragments in a 28.35-gram (one-ounce) chocolate bar?

c. Suppose you consume seven one-ounce (28.35-gram) bars this week. What is the probability of no insect fragments?

4.7.10 WP The time between arrivals of small aircraft at a county airport is exponentially distributed with a mean of one hour.

a. What is the probability that more than three aircraft arrive within an hour?

b. If 30 separate one-hour intervals are chosen, what is the probability that no interval contains more than three arrivals?

c. Determine the length of an interval of time (in hours) such that the probability that no arrivals occur during the interval is 0.10.

d. What is the probability that the first occurrence in which the time between two arrivals exceeds 1.5 hours is at the third arrival?

4.7.11 The time between calls to a corporate office is exponentially distributed with a mean of 10 minutes.

a. What is the probability that there are more than three calls in one-half hour?

b. What is the probability that there are no calls within one-half hour?

c. Determine x such that the probability that there are no calls within x hours is 0.01.

d. What is the probability that there are no calls within a two-hour interval?

e. If four nonoverlapping one-half-hour intervals are selected, what is the probability that none of these intervals contains any call?

f. What is the probability that the first occurrence in which the time between two calls exceeds 10 minutes is at the fifth call?

4.7.12 WP Derive the formula for the mean and variance of an exponential random variable.

4.7.13 SS If the random variable X has an exponential distribution with mean θ, determine the following:

a. $P(X > \theta)$ **b.** $P(X > 2\theta)$ **c.** $P(X > 3\theta)$

d. How do the results depend on θ?

4.7.14 The length of stay at a specific emergency department in a hospital in Phoenix, Arizona, had a mean of 4.6 hours. Assume that the length of stay is exponentially distributed.

a. What is the standard deviation of the length of stay?

b. What is the probability of a length of stay of more than 10 hours?

c. What length of stay is exceeded by 25% of the visits?

4.7.15 An article in the *Journal of the National Cancer Institute* ["Breast Cancer Screening Policies in Developing Countries: A Cost-Effectiveness Analysis for India" (2008, Vol. 100(18), pp. 1290–1300)] presented a screening analysis model of breast cancer based on data from India. In this analysis, the time that a breast cancer case stays in a preclinical state is modeled to be exponentially distributed with a mean depending on the state. For example, the time that a cancer case stays in the state of T1C (tumor size of 11–20 mm) is exponentially distributed with a mean of 1.48 years.

a. What is the probability that a breast cancer case in India stays in the state of T1C for more than 2.0 years?

b. What is the proportion of breast cancer cases in India that spend at least 1.0 year in the state of T1C?

c. Assume that a person in India is diagnosed to be in the state of T1C. What is the probability that the patient is in the same state six months later?

4.7.16 An article in *Ad Hoc Networks* ["Underwater Acoustic Sensor Networks: Target Size Detection and Performance Analysis" (2009, Vol. 7(4), pp. 803–808)] discussed an underwater acoustic sensor network to monitor a given area in an ocean. The network does not use cables and does not interfere with shipping activities. The arrival of clusters of signals generated by the same pulse is taken as a Poisson arrival process with a mean of λ per unit time. Suppose that for a specific underwater acoustic sensor network, this Poisson process has a rate of 2.5 arrivals per unit time.

a. What is the mean time between 2.0 consecutive arrivals?

b. What is the probability that there are no arrivals within 0.3 time units?

c. What is the probability that the time until the first arrival exceeds 1.0 unit of time?

d. Determine the mean arrival rate such that the probability is 0.9 that there are no arrivals in 0.3 time units.

4.7.17 An article in *Vaccine* ["Modeling the Effects of Influenza Vaccination of Health Care Workers in Hospital Departments" (2009, Vol. 27(44), pp. 6261–6267)] considered the immunization of healthcare workers to reduce the hazard rate of influenza

virus infection for patients in regular hospital departments. In this analysis, each patient's length of stay in the department is taken as exponentially distributed with a mean of 7.0 days.

a. What is the probability that a patient stays in a hospital for less than 5.5 days?

b. What is the probability that a patient stays in a hospital for more than 10.0 days if the patient has currently stayed for 7.0 days?

c. Determine the mean length of stay such that the probability is 0.9 that a patient stays in a hospital less than 6.0 days.

Exercises for Section 4.8

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

4.8.1 **WP** **SS** Use the properties of the gamma function to evaluate the following:

a. $\Gamma(6)$ **b.** $\Gamma(5/2)$ **c.** $\Gamma(9/2)$

4.8.2 **WP** Given the probability density function $f(x) = 0.01^3 x^2 e^{-0.01x}/\Gamma(3)$, determine the mean and variance of the distribution.

4.8.3 **WP** Calls to a telephone system follow a Poisson process with a mean of five calls per minute.

a. What is the name applied to the distribution and parameter values of the time until the 10th call?

b. What is the mean time until the 10th call?

c. What is the mean time between the 9th and 10th calls?

d. What is the probability that exactly four calls occur within 1 minute?

e. If 10 separate 1-minute intervals are chosen, what is the probability that all intervals contain more than two calls?

4.8.4 **WP** Raw materials are studied for contamination. Suppose that the number of particles of contamination per pound of material is a Poisson random variable with a mean of 0.01 particle per pound.

a. What is the expected number of pounds of material required to obtain 15 particles of contamination?

b. What is the standard deviation of the pounds of materials required to obtain 15 particles of contamination?

4.8.5 **WP** The time between arrivals of customers at an automatic teller machine is an exponential random variable with a mean of 5 minutes.

a. What is the probability that more than three customers arrive in 10 minutes?

b. What is the probability that the time until the fifth customer arrives is less than 15 minutes?

4.8.6 **WP** In a data communication system, several messages that arrive at a node are bundled into a packet before they are transmitted over the network. Assume that the messages arrive at the node according to a Poisson process with $\lambda = 30$ messages per minute. Five messages are used to form a packet.

a. What is the mean time until a packet is formed, that is, until five messages have arrived at the node?

b. What is the standard deviation of the time until a packet is formed?

c. What is the probability that a packet is formed in less than 10 seconds?

d. What is the probability that a packet is formed in less than 5 seconds?

4.8.7 Patients arrive at a hospital emergency department according to a Poisson process with a mean of 6.5 per hour.

a. What is the mean time until the 10th arrival?

b. What is the probability that more than 20 minutes is required for the third arrival?

4.8.8 Use integration by parts to show that $\Gamma(r) = (r-1)\,\Gamma(r-1)$.

4.8.9 **SS** An article in *Sensors and Actuators A: Physical* ["Characterization and Simulation of Avalanche PhotoDiodes for Next-Generation Colliders" (2011, Vol. 172(1), pp. 181–188)] considered an avalanche photodiode (APD) to detect charged particles in a photo. The number of arrivals in each detection window was modeled with a Poisson distribution with a mean depending on the intensity of beam. For one beam intensity, the number of electrons arriving at an APD follows a Poisson distribution with a mean of 1.74 particles per detection window of 200 nanoseconds.

a. What is the mean and variance of the time for 100 arrivals?

b. What is the probability that the time until the fifth particle arrives is greater than 1.0 nanosecond?

4.8.10 The total service time of a multistep manufacturing operation has a gamma distribution with mean 18 minutes and standard deviation 6.

a. Determine the parameters λ and r of the distribution.

b. Assume that each step has the same distribution for service time. What distribution for each step and how many steps produce this gamma distribution of total service time?

4.8.11 An article in *Mathematical Biosciences* ["Influence of Delayed Viral Production on Viral Dynamics in HIV-1 Infected Patients" (1998, Vol. 152(2), pp. 143–163)] considered the time delay between the initial infection by immunodeficiency virus type 1 (HIV-1) and the formation of productively infected cells. In the simulation model, the time delay is approximated by a gamma distribution with parameters $r = 4$ and $1/\lambda = 0.25$ days. Determine the following:

a. Mean and variance of time delay

b. Probability that a time delay is more than half a day

c. Probability that a time delay is between one-half and one day

Exercises for Section 4.9

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

4.9.1 **WP** If X is a Weibull random variable with $\beta = 1$ and $\delta = 1000$, what is another name for the distribution of X, and what is the mean of X?

4.9.2 **WP** Suppose that X has a Weibull distribution with $\beta = 0.2$ and $\delta = 100$ hours. Determine the following:

a. $P(X < 10{,}000)$ **b.** $P(X > 5000)$
c. $E(X)$ and $V(X)$

4.9.3 **WP** **SS** The life (in hours) of a magnetic resonance imaging machine (MRI) is modeled by a Weibull distribution with parameters $\beta = 2$ and $\delta = 500$ hours. Determine the following:

a. Mean life of the MRI
b. Variance of the life of the MRI
c. Probability that the MRI fails before 250 hours.

4.9.4 **WP** **VS** Assume that the life of a roller bearing follows a Weibull distribution with parameters $\beta = 2$ and $\delta = 10{,}000$ hours.

a. Determine the probability that a bearing lasts at least 8000 hours.
b. Determine the mean time until failure of a bearing.
c. If 10 bearings are in use and failures occur independently, what is the probability that all 10 bearings last at least 8000 hours?

4.9.5 An article in the *Journal of Geophysical Research* ["Spatial and Temporal Distributions of U.S. Winds and Wind Power at 80 m Derived from Measurements" (2003, Vol. 108)] considered wind speed at stations throughout the United States. For a station at Amarillo, Texas, the mean wind speed at 80 m (the hub height of large wind turbines) was 10.3 m/s with a standard deviation of 4.9 m/s. Determine the shape and scale parameters of a Weibull distribution with these properties.

4.9.6 **WP** An article in the *Journal of the Indian Geophysical Union* titled "Weibull and Gamma Distributions for Wave Parameter Predictions" (2005, Vol. 9, pp. 55–64) described the use of the Weibull distribution to model ocean wave heights. Assume that the mean wave height at the observation station is 2.5 m and the shape parameter equals 2. Determine the standard deviation of wave height.

4.9.7 Suppose that the lifetime of a component (in hours) is modeled with a Weibull distribution with $\beta = 2$ and $\delta = 4000$. Determine the following in parts (a) and (b):

a. $P(X > 5000)$
b. $P(X > 8000 \mid X > 3000)$
c. Comment on the probabilities in the previous parts compared to the results for an exponential distribution.

4.9.8 Suppose that X has a Weibull distribution with $\beta = 2$ and $\delta = 8.6$. Determine the following:

a. $P(X < 10)$ **b.** $P(X > 9)$ **c.** $P(8 < X < 11)$
d. Value for x such that $P(X > x) = 0.9$

4.9.9 **SS** An article in *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval* ["Understanding Web Browsing Behaviors Through Weibull Analysis of Dwell Time" (2010, pp. 379–386)] proposed that a Weibull distribution can be used to model Web page dwell time (the length of time a Web visitor spends on a Web page). For a specific Web page, the shape and scale parameters are 1 and 300 seconds, respectively. Determine the following:

a. Mean and variance of dwell time
b. Probability that a Web user spends more than 4 minutes on this Web page
c. Dwell time exceeded with probability 0.25

4.9.10 An article in *Electronic Journal of Applied Statistical Analysis* ["Survival Analysis of Dialysis Patients Under Parametric and Non-Parametric Approaches" (2012, Vol. 5(2), pp. 271–288)] modeled the survival time of dialysis patients with chronic kidney disease with a Weibull distribution. The mean and standard deviation of survival time were 16.01 and 11.66 months, respectively. Determine the following:

a. Shape and scale parameters of this Weibull distribution
b. Probability that survival time is more than 48 months
c. Survival time exceeded with 90% probability

4.9.11 An article in *IEEE Transactions on Dielectrics and Electrical Insulation* ["Statistical Analysis of the AC Breakdown Voltages of Ester Based Transformer Oils" (2008, Vol. 15(4), pp. 1044–1050)] used Weibull distributions to model the breakdown voltage of insulators. The breakdown voltage is the minimum voltage at which the insulator conducts. For 1 mm of natural ester, the 1% probability of breakdown voltage is approximately 26 kV, and the 7% probability is approximately 31.6 kV. Determine the parameters δ and β of the Weibull distribution.

4.9.12 An article in *Financial Markets Institutions and Instruments* ["Pricing Reinsurance Contracts on FDIC Losses" (2008, Vol. 17(3), pp. 225–247)] modeled average annual losses (in billions of dollars) of the Federal Deposit Insurance Corporation (FDIC) with a Weibull distribution with parameters $\delta = 1.9317$ and $\beta = 0.8472$. Determine the following:

a. Probability of a loss greater than \$2 billion
b. Probability of a loss between \$2 and \$4 billion
c. Value exceeded with probability 0.05
d. Mean and standard deviation of loss

Exercises for Section 4.10

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

4.10.1 WP Suppose that X has a lognormal distribution with parameters $\theta = 5$ and $\omega^2 = 9$. Determine the following:

- **a.** $P(X < 13{,}300)$
- **b.** Value for x such that $P(X \leq x) = 0.95$
- **c.** Mean and variance of X

4.10.2 WP Suppose that X has a lognormal distribution with parameters $\theta = 2$ and $\omega^2 = 4$. Determine the following in parts (a) and (b):

- **a.** $P(X < 500)$
- **b.** Conditional probability that $X < 1500$ given that $X > 1000$
- **c.** What does the difference between the probabilities in parts (a) and (b) imply about lifetimes of lognormal random variables?

4.10.3 WP SS VS The length of time (in seconds) that a user views a page on a Web site before moving to another page is a lognormal random variable with parameters $\theta = 0.5$ and $\omega^2 = 1$.

- **a.** What is the probability that a page is viewed for more than 10 seconds?
- **b.** By what length of time have 50% of the users moved to another page?
- **c.** What are the mean and standard deviation of the time until a user moves from the page?

4.10.4 WP Suppose that X has a lognormal distribution and that the mean and variance of X are 100 and 85,000, respectively. Determine the parameters θ and ω^2 of the lognormal distribution. [*Hint*: Define $u = \exp(\theta)$ and $v = \exp(\omega^2)$ and write two equations in terms of u and v.]

4.10.5 WP The lifetime of a semiconductor laser has a lognormal distribution, and it is known that the mean and standard deviation of lifetime are 10,000 and 20,000, respectively.

- **a.** Calculate the parameters of the lognormal distribution.
- **b.** Determine the probability that a lifetime exceeds 10,000 hours.
- **c.** Determine the lifetime that is exceeded by 90% of lasers.

4.10.6 WP An article in *Health and Population: Perspectives and Issues* (2000, Vol. 23, pp. 28–36) used the lognormal distribution to model blood pressure in humans. The mean systolic blood pressure (SBP) in males age 17 was 120.87 mm Hg. If the coefficient of variation (100% × Standard deviation/mean) is 9%, what are the parameter values of the lognormal distribution?

4.10.7 Suppose that the length of stay (in hours) at a hospital emergency department is modeled with a lognormal random variable X with $\theta = 1.5$ and $\omega = 0.4$. Determine the following in parts (a) and (b):

- **a.** Mean and variance
- **b.** $P(X < 8)$
- **c.** Comment on the difference between the probability $P(X < 0)$ calculated from this lognormal distribution and a normal distribution with the same mean and variance.

4.10.8 An article in *Journal of Hydrology* ["Use of a Lognormal Distribution Model for Estimating Soil Water Retention Curves from Particle-Size Distribution Data" (2006, Vol. 323(1), pp. 325–334)] considered a lognormal distribution model to estimate water retention curves for a range of soil textures. The particle-size distribution (in centimeters) was modeled as a lognormal random variable X with $\theta = -3.8$ and $\omega = 0.7$. Determine the following:

- **a.** $P(X < 0.02)$
- **b.** Value for x such that $P(X \leq x) = 0.95$
- **c.** Mean and variance of X

4.10.9 SS An article in *Applied Mathematics and Computation* ["Confidence Intervals for Steady State Availability of a System with Exponential Operating Time and Lognormal Repair Time" (2003, Vol. 137(2), pp. 499–509)] considered the long-run availability of a system with an assumed lognormal distribution for repair time. In a given example, repair time follows a lognormal distribution with $\theta = \omega = 1$. Determine the following:

- **a.** Probability that repair time is greater than five time units
- **b.** Conditional probability that a repair time is less than eight time units given that it is more than five time units
- **c.** Mean and variance of repair time

4.10.10 An article in *Chemosphere* ["Statistical Evaluations Reflecting the Skewness in the Distribution of TCDD Levels in Human Adipose Tissue" (1987, Vol. 16(8), pp. 2135–2140)] concluded that the levels of 2,3,7,8-TCDD (colorless persistent environmental contaminants with no distinguishable odor at room temperature) in human adipose tissue has a lognormal distribution. The mean and variance of this lognormal distribution in the United States are 8 and 21, respectively. Let X denote this lognormal random variable. Determine the following:

- **a.** $P(2000 < X < 2500)$
- **b.** Value exceeded with probability 10%
- **c.** Mean and variance of X

Exercises for Section 4.11

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

4.11.1 Suppose that x has a beta distribution with parameters $\alpha = 2.5$ and $\beta = 1$. Determine the following:

a. $P(X < 0.25)$ **b.** $P(0.25 < X < 0.75)$

c. Mean and variance

4.11.2 Suppose that X has a beta distribution with parameters $\alpha = 2.5$ and $\beta = 2.5$. Sketch an approximate graph of the probability density function. Is the density symmetric?

4.11.3 A European standard value for a low-emission window glazing uses 0.59 as the proportion of solar energy that enters a room. Suppose that the distribution of the proportion of solar energy that enters a room is a beta random variable.

a. Calculate the mode, mean, and variance of the distribution for $\alpha = 3$ and $\beta = 1.4$.

b. Calculate the mode, mean, and variance of the distribution for $\alpha = 10$ and $\beta = 6.25$.

c. Comment on the difference in dispersion in the distribution from parts (a) and (b).

4.11.4 The length of stay at a hospital emergency department is the sum of the waiting and service times. Let X denote the proportion of time spent waiting and assume a beta distribution with $\alpha = 10$ and $\beta = 1$. Determine the following:

a. $P(X > 0.9)$ **b.** $P(X < 0.5)$

c. Mean and variance

4.11.5 SS The maximum time to complete a task in a project is 2.5 days. Suppose that the completion time as a proportion of this maximum is a beta random variable with $\alpha = 2$ and $\beta = 3$. What is the probability that the task requires more than two days to complete?

4.11.6 An *allele* is an alternate form of a gene, and the proportion of alleles in a population is of interest in genetics. An article in *BMC Genetics* ["Calculating Expected DNA Remnants from Ancient Founding Events in Human Population Genetics" (2008, Vol. 9, p. 66)] used a beta distribution with mean 0.3 and standard deviation 0.17 to model initial allele proportions in a genetic simulation. Determine the parameters α and β for this beta distribution.

4.11.7 Suppose that the construction of a solar power station is initiated. The project's completion time has not been set due to uncertainties in financial resources. The completion time for the first phase is modeled with a beta distribution and the minimum, most likely (mode), and maximum completion times for the first phase are 1.0, 1.25, and 2.0 years, respectively. Also, the mean time is assumed to equal $\mu = 1 + 4(1.25) + 2/6 = 1.333$. Determine the following in parts (a) and (b):

a. Parameters α and β of the beta distribution.

b. Standard deviation of the distribution.

c. Sketch the probability density function.

Supplemental Exercises for Chapter 4

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

4.S8 WP The probability density function of the time it takes a hematology cell counter to complete a test on a blood sample is $f(x) = 0.04$ for $50 < x < 75$ seconds.

a. What percentage of tests requires more than 70 seconds to complete?

b. What percentage of tests requires less than one minute to complete?

c. Determine the mean and variance of the time to complete a test on a sample.

4.S9 WP The time it takes a cell to divide (called *mitosis*) is normally distributed with an average time of 1 hour and a standard deviation of 5 minutes.

a. What is the probability that a cell divides in less than 45 minutes?

b. What is the probability that it takes a cell more than 65 minutes to divide?

c. By what time have approximately 99% of all cells completed mitosis?

4.S10 WP The sick-leave time of employees in a firm in a month is normally distributed with a mean of 100 hours and a standard deviation of 20 hours.

a. What is the probability that the sick-leave time for next month will be between 50 and 80 hours?

b. How much time should be budgeted for sick leave if the budgeted amount should be exceeded with a probability of only 10%?

4.S11 SS The percentage of people exposed to a bacteria who become ill is 20%. Assume that people are independent. Assume that 1000 people are exposed to the bacteria. Approximate each of the following:

a. Probability that more than 225 become ill

b. Probability that between 175 and 225 become ill

c. Value such that the probability that the number of people who become ill exceeds the value is 0.01

4.S12 WP The time to failure (in hours) for a laser in a cytometry machine is modeled by an exponential distribution with $\lambda = 0.00004$. What is the probability that the time until failure is

a. At least 20,000 hours?

b. At most 30,000 hours?

c. Between 20,000 and 30,000 hours?

4.S13 WP When a bus service reduces fares, a particular trip from New York City to Albany, New York, is very popular. A small bus can carry four passengers. The time between calls for tickets is exponentially distributed with a mean of 30 minutes. Assume that each caller orders one ticket. What is the probability that the bus is filled in less than three hours from the time of the fare reduction?

4.S14 The size of silver particles in a photographic emulsion is known to have a log normal distribution with a mean of 0.001 mm and a standard deviation of 0.002 mm.

a. Determine the parameter values for the lognormal distribution.

b. What is the probability of a particle size greater than 0.005 mm?

4.S15 The life of a recirculating pump follows a Weibull distribution with parameters $\beta = 2$ and $\delta = 700$ hours. Determine for parts (a) and (b):

a. Mean life of a pump

b. Variance of the life of a pump

c. What is the probability that a pump will last longer than its mean?

4.S16 WP The time between calls is exponentially distributed with a mean time between calls of 10 minutes.

a. What is the probability that the time until the first call is less than five minutes?

b. What is the probability that the time until the first call is between 5 and 15 minutes?

c. Determine the length of an interval of time such that the probability of at least one call in the interval is 0.90.

d. If there has not been a call in 10 minutes, what is the probability that the time until the next call is less than 5 minutes?

e. What is the probability that there are no calls in the intervals from 10:00 to 10:05, from 11:30 to 11:35, and from 2:00 to 2:05?

f. What is the probability that the time until the third call is greater than 30 minutes?

g. What is the mean time until the fifth call?

4.S17 Suppose that $f(x) = 0.5x - 1$ for $2 < x < 4$. Determine the following:

a. $P(X < 2.5)$ **b.** $P(X > 3)$ **c.** $P(2.5 < X < 3.5)$

d. Determine the cumulative distribution function of the random variable.

e. Determine the mean and variance of the random variable.

4.S18 WP Suppose that X has a lognormal distribution with parameters $\theta = 0$ and $\omega^2 = 4$. Determine the following:

a. $P(10 < X < 50)$

b. Value for x such that $P(X < x) = 0.05$

c. Mean and variance of X

4.S19 WP SS Suppose that X has a lognormal distribution and that the mean and variance of X are 50 and 4000, respectively. Determine the following:

a. Parameters θ and ω^2 of the lognormal distribution

b. Probability that X is less than 150

4.S20 Asbestos fibers in a dust sample are identified by an electron microscope after sample preparation. Suppose that the number of fibers is a Poisson random variable and the mean number of fibers per square centimeter of surface dust is 100. A sample of 800 square centimeters of dust is analyzed. Assume that a particular grid cell under the microscope represents 1/160,000 of the sample.

a. What is the probability that at least one fiber is visible in the grid cell?

b. What is the mean of the number of grid cells that need to be viewed to observe 10 that contain fibers?

c. What is the standard deviation of the number of grid cells that need to be viewed to observe 10 that contain fibers?

4.S21 The waiting time for service at a hospital emergency department follows an exponential distribution with a mean of three hours. Determine the following:

a. Waiting time is greater than four hours

b. Waiting time is greater than six hours given that you have already waited two hours

c. Value x (in hours) exceeded with probability 0.25

4.S22 WP The diameter of the dot produced by a printer is normally distributed with a mean diameter of 0.002 inch.

a. Suppose that the specifications require the dot diameter to be between 0.0014 and 0.0026 inch. If the probability that a dot meets specifications is to be 0.9973, what standard deviation is needed?

b. Assume that the standard deviation of the size of a dot is 0.0004 inch. If the probability that a dot meets specifications is to be 0.9973, what specifications are needed? Assume that the specifications are to be chosen symmetrically around the mean of 0.002.

4.S23 WP A square inch of carpeting contains 50 carpet fibers. The probability of a damaged fiber is 0.0001. Assume that the damaged fibers occur independently.

a. Approximate the probability of one or more damaged fibers in one square yard of carpeting.

b. Approximate the probability of four or more damaged fibers in one square yard of carpeting.

4.S24 An airline makes 200 reservations for a flight that holds 185 passengers. The probability that a passenger arrives for the flight is 0.9, and the passengers are assumed to be independent.

a. Approximate the probability that all the passengers who arrive can be seated.

b. Approximate the probability that the flight has empty seats.

c. Approximate the number of reservations that the airline should allow so that the probability that everyone who arrives can be seated is 0.95. [*Hint*: Successively try values for the number of reservations.]

4.S25 Suppose that the construction of a solar power station is initiated. The project's completion time has not been set due to uncertainties in financial resources. The proportion of

completion within one year has a beta distribution with parameters $\alpha = 1$ and $\beta = 5$. Determine the following:

a. Mean and variance of the proportion completed within one year

b. Probability that more than half of the project is completed within one year

c. Proportion of the project that is completed within one year with probability 0.9

4.S26 An article in *IEEE Journal on Selected Areas in Communications* ["Impulse Response Modeling of Indoor Radio Propagation Channels" (1993, Vol. 11(7), pp. 967–978)] indicated that the successful design of indoor communication systems requires characterization of radio propagation. The distribution of the amplitude of individual multipath components was well modeled with a lognormal distribution. For one test configuration (with 100 ns delays), the mean amplitude was −24 dB (from the peak) with a standard deviation of 4.1 dB. The amplitude decreased nearly linearly with increased excess delay. Determine the following:

a. Probability the amplitude exceeds −20 dB

b. Amplitude exceeded with probability 0.05

4.S27 SS Consider the regional right ventricle transverse wall motion in patients with pulmonary hypertension (PH). The right-ventricle ejection fraction (EF) is approximately normally distributed with a standard deviation of 12 for PH subjects, and with mean and standard deviation of 56 and 8, respectively, for control subjects.

a. What is the EF for control subjects exceeded with 99% probability?

b. What is the mean for PH subjects such that the probability is 1% that the EF of a PH subject is greater than the value in part (a)?

c. Comment on how well the control and PH subjects [with the mean determined in part (b)] can be distinguished by EF measurements.

4.S28 Provide approximate sketches for beta probability density functions with the following parameters. Comment on any symmetries and show any peaks in the probability density functions in the sketches.

a. $\alpha = \beta < 1$ **b.** $\alpha = \beta = 1$ **c.** $\alpha = \beta > 1$

4.S29 Among homeowners in a metropolitan area, 25% recycle paper each week. A waste management company services 10,000 homeowners (assumed independent). Approximate the following probabilities:

a. More than 2600 recycle paper in a week

b. Between 2400 and 2600 recycle paper in a week

c. Number of customers who recycle paper in a week that is exceeded with probability approximately 0.05

4.S30 An article in *Journal of Theoretical Biology* ["Computer Model of Growth Cone Behavior and Neuronal Morphogenesis" (1995, Vol. 174(4), pp. 381–389)] developed a model for neuronal morphogenesis in which neuronal growth cones have a significant function in the development of the nervous system. This model assumes that the time interval between filopodium formation (a process in growth cone behavior) is exponentially distributed with a mean of 6 time units. Determine the following:

a. Probability formation requires more than nine time units

b. Probability formation occurs within six to seven time units

c. Formation time exceeded with probability 0.9

4.S31 An article in *Electric Power Systems Research* ["On the Self-Scheduling of a Power Producer in Uncertain Trading Environments" (2008, Vol. 78(3), pp. 311–317)] considered a self-scheduling approach for a power producer. In addition to price and forced outages, another uncertainty was due to generation reallocations to manage congestions. Generation reallocation was modeled as $110X - 60$ (with range [−60,50] MW/h) where X has a beta distribution with parameters $\alpha = 3.2$ and $\beta = 2.8$. Determine the mean and variance of generation reallocation.

4.S32 An article in *Electronic Journal of Applied Statistical Analysis* ["Survival Analysis of Acute Myocardial Infarction Patients Using Non-Parametric and Parametric Approaches" (2009, Vol. 2(1), pp. 22–36)] described the use of a Weibull distribution to model the survival time of acute myocardial infarction (AMI) patients in a hospital-based retrospective study. The shape and scale parameters for the Weibull distribution in the model were 1.16 and 0.25 years, respectively. Determine the following:

a. Mean and standard deviation of survival time

b. Probability that a patient survives more than a year

c. Survival time exceeded with probability 0.9

Chapter 5 Exercises

Exercises for Section 5.1

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

5.1.1 WP Show that the following function satisfies the properties of a joint probability mass function.

x	y	$f_{XY}(x, y)$
1.0	1	1/4
1.5	2	1/8
1.5	3	1/4
2.5	4	1/4
3.0	5	1/8

Determine the following:

a. $P(X < 2.5, Y < 3)$ **b.** $P(X < 2.5)$
c. $P(Y < 3)$ **d.** $P(X > 1.8, Y > 4.7)$
e. $E(X), E(Y), V(X), V(Y)$
f. Marginal probability distribution of X

5.1.2 WP Show that the following function satisfies the properties of a joint probability mass function.

x	y	$f_{XY}(x, y)$
−1.0	−2	1/8
−0.5	−1	1/4
0.5	1	1/2
1.0	2	1/8

Determine the following:

a. $P(X < 0.5, Y < 1.5)$ **b.** $P(X < 0.5)$
c. $P(Y < 1.5)$ **d.** $P(X > 0.25, Y < 4.5)$
e. $E(X), E(Y), V(X), V(Y)$
f. Marginal probability distribution of X

5.1.3 WP SS VS In the transmission of digital information, the probability that a bit has high, moderate, and low distortion is 0.01, 0.04, and 0.95, respectively. Suppose that three bits are transmitted and that the amount of distortion of each bit is assumed to be independent. Let X and Y denote the number of bits with high and moderate distortion out of the three, respectively. Determine:

a. $f_{XY}(x, y)$ **b.** $f_X(x)$ **c.** $E(X)$

5.1.4 WP A small-business Web site contains 100 pages and 60%, 30%, and 10% of the pages contain low, moderate, and high graphic content, respectively. A sample of four pages is selected randomly without replacement, and X and Y denote the number of pages in the sample with moderate and high graphics output. Determine:

a. $f_{XY}(x, y)$ **b.** $f_X(x)$ **c.** $E(X)$

5.1.5 WP A manufacturing company employs two devices to inspect output for quality control purposes. The first device is able to accurately detect 99.3% of the defective items it receives, whereas the second is able to do so in 99.7% of the cases. Assume that four defective items are produced and sent out for inspection. Let X and Y denote the number of items that will be identified as defective by inspecting devices 1 and 2, respectively. Assume that the devices are independent. Determine:

a. $f_{XY}(x, y)$ **b.** $f_X(x)$ **c.** $E(X)$

5.1.6 WP An article in the *Journal of Database Management* ["Experimental Study of a Self-Tuning Algorithm for DBMS Buffer Pools" (2005, Vol. 16, pp. 1–20)] provided the workload used in the TPC-C OLTP (Transaction Processing Performance Council's Version C On-Line Transaction Processing) benchmark, which simulates a typical order entry application. See the table at the bottom of the page. The frequency of each type of transaction (in the second column) can be used as the percentage of each type of transaction. Let X and Y denote the average number of *selects* and *updates* operations, respectively, required for each type transaction. Determine the following:

a. $P(X < 5)$ **b.** $E(X)$

5.1.7 An engineering statistics class has 40 students; 60% are electrical engineering majors, 10% are industrial engineering majors, and 30% are mechanical engineering majors. A sample of four students is selected randomly without replacement for a project team. Let X and Y denote the number of industrial and mechanical engineering majors in the sample, respectively. Determine the following:

a. $f_{XY}(x, y)$ **b.** $f_X(x)$ **c.** $E(X)$

5.1.8 WP Determine the value of c that makes the function $f(x, y) = c(x + y)$ a joint probability density function over the range $0 < x < 3$ and $x < y < x + 2$. Determine the following:

a. $P(X < 1, Y < 2)$ **b.** $P(1 < X < 2)$ **c.** $P(Y > 1)$
d. $P(X < 2, Y < 2)$ **e.** $E(X)$ **f.** $V(X)$
g. Marginal probability distribution of X

5.1.9 WP SS VS Determine the value of c such that the function $f(x, y) = cxy$ for $0 < x < 3$ and $0 < y < 3$ satisfies the properties of a joint probability density function. Determine the following:

a. $P(X < 2, Y < 3)$ **b.** $P(X < 2.5)$
c. $P(1 < Y < 2.5)$ **d.** $P(X > 1.8, 1 < Y < 2.5)$
e. $E(X)$ **f.** $P(X < 0, Y < 4)$
g. Marginal probability distribution of X

5.1.10 VS Determine the value of c that makes the function $f(x, y) = ce^{-2x-3y}$ a joint probability density function over the range $0 < x$ and $x < y$. Determine the following:

a. $P(X < 1, Y < 2)$ **b.** $P(1 < X < 2)$
c. $P(Y > 3)$ **d.** $P(X < 2, Y < 2)$
e. $E(X)$ **f.** $E(Y)$
g. Marginal probability distribution of X

Average Frequencies and Operations in TPC-C

Transaction	Frequency	Selects	Updates	Inserts	Deletes	Non-unique Selects	Joins
New order	43	23.0	11	12	0	0	0
Payment	44	4.2	3	1	0	0.6	0
Order status	4	11.4	0	0	0	0.6	0
Delivery	5	130.0	120	0	10	0	0
Stock level	4	0	0	0	0	0	1

Exercises for Section 5.2

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

5.2.1 The conditional probability distribution of Y given $X = x$ is $f_{Y|x}(y) = xe^{-xy}$ for $y > 0$, and the marginal probability distribution of X is a continuous uniform distribution over 0 to 10.

a. Graph $f_{Y|x}(y) = xe^{-xy}$ for $y > 0$ for several values of x. Determine:

b. $P(Y < 2 \mid X = 2)$ **c.** $E(Y \mid X = 2)$
d. $E(Y \mid X = x)$ **e.** $f_{XY}(x, y)$
f. $f_Y(y)$

5.2.2 **WP** Consider the joint distribution in Exercise 5.1.1. Determine the following:

a. Conditional probability distribution of Y given that $X = 1.5$
b. Conditional probability distribution of X that $Y = 2$
c. $E(Y \mid X = 1.5)$
d. Are X and Y independent?

5.2.3 **WP** Consider the joint distribution in Exercise 5.1.4. Determine the following:

a. $f_{Y|3}(y)$ **b.** $E(Y \mid X = 3)$ **c.** $V(Y \mid X = 3)$

5.2.4 **WP** A manufacturer of electroluminescent lamps knows that the amount of luminescent ink deposited on one of its products is normally distributed with a mean of 1.2 grams and a standard deviation of 0.03 gram. Any lamp with less than 1.14 grams of luminescent ink fails to meet customers' specifications. A random sample of 25 lamps is collected and the mass of luminescent ink on each is measured.

a. What is the probability that at least one lamp fails to meet specifications?
b. What is the probability that five or fewer lamps fail to meet specifications?
c. What is the probability that all lamps conform to specifications?
d. Why is the joint probability distribution of the 25 lamps not needed to answer the previous questions?

5.2.5 **SS** The lengths of the minor and major axes are used to summarize dust particles that are approximately elliptical in shape. Let X and Y denote the lengths of the minor and major axes (in micrometers), respectively. Suppose that $f_X(x) = \exp(-x), 0 < x$ and the conditional distribution $f_{Y|x}(y) = \exp[-(y - x)], x < y$. Answer or determine the following:

a. Show that $f_{Y|x}(y)$ is a probability density function for any value of x
b. $P(X < Y)$
c. Joint probability density function $f_{XY}(x, y)$
d. Conditional probability density function of X given $Y = y$
e. $P(Y < 2 \mid X = 1)$ **f.** $E(Y \mid X = 1)$
g. $P(X < 1, Y < 1)$ **h.** $P(Y < 2)$
i. c such that $P(Y < c) = 0.9$
j. Are X and Y independent?

5.2.6 An article in *Health Economics* ["Estimation of the Transition Matrix of a Discrete-Time Markov Chain" (2002, Vol. 11, pp. 33–42)] considered the changes in CD4 white blood cell counts from one month to the next. The CD4 count is an important clinical measure to determine the severity of HIV infections. The CD4 count was grouped into three distinct categories: 0–49, 50–74, and ≥ 75. Let X and Y denote the (category minimum) CD4 count at a month and the following month, respectively. The conditional probabilities for Y given values for X were provided by a transition probability matrix shown in the following table.

X	Y		
	0	50	75
0	0.9819	0.0122	0.0059
50	0.1766	0.7517	0.0717
75	0.0237	0.0933	0.8830

This table is interpreted as follows. For example, $P(Y = 50 \mid X = 75) = 0.0933$. Suppose also that the probability distribution for X is $P(X = 75) = 0.9, P(X = 50) = 0.08, P(X = 0) = 0.02$. Determine the following:

a. $P(Y \leq 50 \mid X = 50)$
b. $P(X = 0, Y = 75)$
c. $E(Y \mid X = 50)$ **d.** $f_Y(y)$
e. $f_{XY}(x, y)$ **f.** Are X and Y independent?

5.2.7 An article in *Clinical Infectious Diseases* ["Strengthening the Supply of Routinely Administered Vaccines in the United States: Problems and Proposed Solutions" (2006, Vol. 42(3), pp. S97–S103)] reported that recommended vaccines for infants and children were periodically unavailable or in short supply in the United States. Although the number of doses demanded each month is a discrete random variable, the large demands can be approximated with a continuous probability distribution. Suppose that the monthly demands for two of those vaccines, namely measles–mumps–rubella (MMR) and varicella (for chickenpox), are independently, normally distributed with means of 1.1 and 0.55 million doses and standard deviations of 0.3 and 0.1 million doses, respectively. Also suppose that the inventory levels at the beginning of a given month for MMR and varicella vaccines are 1.2 and 0.6 million doses, respectively.

a. What is the probability that there is no shortage of either vaccine in a month without any vaccine production?
b. To what should inventory levels be set so that the probability is 90% that there is no shortage of either vaccine in a month without production? Can there be more than one answer? Explain.

5.2.8 The systolic and diastolic blood pressure values (mm Hg) are the pressures when the heart muscle contracts and relaxes (denoted as Y and X, respectively). Over a collection of individuals, the distribution of diastolic pressure is normal with mean 73 and standard deviation 8. The systolic pressure is conditionally

normally distributed with mean $1.6x$ when $X = x$ and standard deviation of 10. Determine the following:

a. Conditional probability density function of Y given $X = 73$

b. $P(Y < 115 \mid X = 73)$

c. $E(Y \mid X = 73)$

d. Recognize the distribution $f_{XY}(x, y)$ and identify the mean and variance of Y

5.2.9 WP SS VS Consider the joint distribution in Exercise 5.1.10. Determine the following:

a. $f_{Y|2}(y)$ **b.** $E(Y \mid X = 2)$

c. $V(Y \mid X = 2)$ **d.** Are X and Y independent?

Exercises for Section 5.3

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

5.3.1 WP SS For the Transaction Processing Performance Council's benchmark in Exercise 5.1.6, let X, Y, and Z denote the average number of *selects, updates*, and *inserts* operations required for each type of transaction, respectively. Calculate the following:

a. $f_{XYZ}(x, y, z)$

b. Conditional probability mass function for X and Y given $Z = 0$

c. $P(X < 6, Y < 6 \mid Z = 0)$ **d.** $E(X \mid Y = 0, Z = 0)$

5.3.2 WP Suppose that the random variables X, Y, and Z have the following joint probability distribution.

x	y	z	$f(x, y, z)$
1	1	1	0.05
1	1	2	0.10
1	2	1	0.15
1	2	2	0.20
2	1	1	0.20
2	1	2	0.15
2	2	1	0.10
2	2	2	0.05

Determine the following:

a. $P(X = 2)$ **b.** $P(X = 1, Y = 2)$

c. $P(Z < 1.5)$ **d.** $P(X = 1 \text{ or } Z = 2)$

e. $E(X)$ **f.** $P(X = 1 \mid Y = 1)$

g. $P(X = 1, Y = 1 \mid Z = 2)$ **h.** $P(X = 1 \mid Y = 1, Z = 2)$

i. Conditional probability distribution of X given that $Y = 1$ and $Z = 2$

5.3.3 WP Suppose that the random variables X, Y, and Z have the joint probability density function $f(x, y, z) = 8xyz$ for $0 < x < 1$, $0 < y < 1$, and $0 < z < 1$. Determine the following:

a. $P(X < 0.5)$ **b.** $P(X < 0.5, Y < 0.5)$

c. $P(Z < 2)$ **d.** $P(X < 0.5 \text{ or } Z < 2)$

e. $E(X)$ **f.** $P(X < 0.5 \mid Y = 0.5)$

g. $P(X < 0.5, Y < 0.5 \mid Z = 0.8)$

h. Conditional probability distribution of X given that $Y = 0.5$ and $Z = 0.8$

i. $P(X < 0.5 \mid Y = 0.5, Z = 0.8)$

5.3.4 Suppose that the random variables X, Y, and Z have the joint probability density function $f_{XYZ}(x, y, z) = c$ over the cylinder $x^2 + y^2 < 4$ and $0 < z < 4$. Determine the constant c so that $f_{XYZ}(x, y, z)$ is a probability density function.

Determine the following:

a. $P(X^2 + Y^2 < 2)$ **b.** $P(Z < 2)$

c. $E(X)$ **d.** $P(X < 1 \mid Y = 1)$

e. $P(X^2 + Y^2 < 1 \mid Z = 1)$

f. Conditional probability distribution of Z given that $X = 1$ and $Y = 1$.

Exercises for Section 5.4

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

5.4.1 WP SS VS Determine the covariance and correlation for the following joint probability distribution:

x	1	1	2	4
y	3	4	5	6
$f_{XY}(x, y)$	1/8	1/4	1/2	1/8

5.4.2 WP Determine the value for c and the covariance and correlation for the joint probability mass function $f_{XY}(x, y) = c(x + y)$ for $x = 1, 2, 3$ and $y = 1, 2, 3$.

5.4.3 WP Patients are given a drug treatment and then evaluated. Symptoms either improve, degrade, or remain the same with probabilities 0.4, 0.1, 0.5, respectively. Assume that four independent patients are treated and let X and Y denote the number of patients who improve and degrade, respectively. Are X and Y independent? Calculate the covariance and correlation between X and Y.

5.4.4 For the Transaction Processing Performance Council's benchmark in Exercise 5.1.6, let X, Y, and Z denote the average number of *selects, updates*, and *inserts* operations required for each type of transaction, respectively. Calculate the following:

a. Correlation between X and Y

b. Correlation between X and Z

5.4.5 WP Determine the value for c and the covariance and correlation for the joint probability density function $f_{XY}(x, y) = cxy$ over the range $0 < x < 3$ and $0 < y < x$.

5.4.6 WP Determine the covariance and correlation for the joint probability density function $f_{XY}(x, y) = e^{-x-y}$ over the range $0 < x$ and $0 < y$.

5.4.7 SS The joint probability distribution is

x	-1	0	0	1
y	0	-1	1	0
$f_{XY}(x, y)$	$1/4$	$1/4$	$1/4$	$1/4$

Show that the correlation between X and Y is zero but X and Y are not independent.

5.4.8 Determine the covariance and correlation for the CD4 counts in a month and the following month in Exercise 5.2.6.

5.4.9 Determine the covariance and correlation for the lengths of the minor and major axes in Exercise 5.2.5.

5.4.10 WP Suppose that the correlation between X and Y is ρ. For constants a, b, c, and d, what is the correlation between the random variables $U = aX + b$ and $V = cY + d$?

Exercises for Section 5.5

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

5.5.1 WP Based on the number of voids, a ferrite slab is classified as either high, medium, or low. Historically, 5% of the slabs are classified as high, 85% as medium, and 10% as low. A group of 20 slabs that are independent regarding voids is selected for testing. Let X, Y, and Z denote the number of slabs that are classified as high, medium, and low, respectively.

a. What are the name and the values of the parameters of the joint probability distribution of X, Y, and Z?

b. What is the range of the joint probability distribution of X, Y, and Z?

c. What are the name and the values of the parameters of the marginal probability distribution of X?

d. Determine $E(X)$ and $V(X)$.

Determine the following:

e. $P(X = 1, Y = 17, Z = 3)$

f. $P(X \leq 1, Y = 17, Z = 3)$

g. $P(X \leq 1)$

h. $E(Y)$

i. $P(X = 2, Z = 3 \mid Y = 17)$

j. $P(X = 2 \mid Y = 17)$

k. $E(X \mid Y = 17)$

5.5.2 A Web site uses ads to route visitors to one of four landing pages. The probabilities for each landing page are equal. Consider 20 independent visitors and let the random variables W, X, Y, and Z denote the number of visitors routed to each page.

Calculate the following:

a. $P(W = 5, X = 5, Y = 5, Z = 5)$

b. $P(W = 5, X = 5, Y = 5, Z = 5)$

c. $P(W = 7, X = 7, Y = 6 \mid Z = 3)$

d. $P(W = 7, X = 7, Y = 3 \mid Z = 3)$

e. $P(W \leq 2)$

f. $E(W)$

g. $P(W = 5, X = 5)$

h. $P(W = 5 \mid X = 5)$

5.5.3 WP Suppose that X and Y have a bivariate normal distribution with $\sigma_X = 0.04$, $\sigma_Y = 0.08$, $\mu_X = 3.00$, $\mu_Y = 7.70$, and $\rho = 0$.

Determine the following:

a. $P(2.95 < X < 3.05)$

b. $P(7.60 < Y < 7.80)$

c. $P(2.95 < X < 3.05, 7.60 < Y < 7.80)$

5.5.4 Let X and Y represent the concentration and viscosity of a chemical product. Suppose that X and Y have a bivariate normal distribution with $\sigma_X = 4$, $\sigma_Y = 1$, $\mu_X = 2$, and $\mu_Y = 1$. Draw a rough contour plot of the joint probability density function for each of the following values of ρ:

a. $\rho = 0$

b. $\rho = 0.8$

c. $\rho = -0.8$

5.5.5 WP SS VS In the manufacture of electroluminescent lamps, several different layers of ink are deposited onto a plastic substrate. The thickness of these layers is critical if specifications regarding the final color and intensity of light are to be met. Let X and Y denote the thickness of two different layers of ink. It is known that X is normally distributed with a mean of 0.1 millimeter and a standard deviation of 0.00031 millimeter, and Y is normally distributed with a mean of 0.23 millimeter and a standard deviation of 0.00017 millimeter. The value of ρ for these variables is equal to 0. Specifications call for a lamp to have a thickness of the ink corresponding to X in the range of 0.099535 to 0.100465 millimeter and Y in the range of 0.22966 to 0.23034 millimeter. What is the probability that a randomly selected lamp will conform to specifications?

5.5.6 WP In an acid-base titration, a base or acid is gradually added to the other until they have completely neutralized each other. Let X and Y denote the milliliters of acid and base needed for equivalence, respectively. Assume that X and Y have

a bivariate normal distribution with $\sigma_X = 5$ mL, $\sigma_Y = 2$ mL, $\mu_X = 120$ mL, $\mu_Y = 100$ mL, and $\rho = 0.6$.
Determine the following:

a. Covariance between X and Y
b. Marginal probability distribution of X
c. $P(X < 116)$
d. Conditional probability distribution of X given that $Y = 102$
e. $P(X < 116 \mid Y = 102)$

5.5.7 If X and Y have a bivariate normal distribution with $\rho = 0$, show that X and Y are independent.

5.5.8 Patients given drug therapy either improve, remain the same, or degrade with probabilities 0.5, 0.4, and 0.1, respectively. Suppose that 20 patients (assumed to be independent) are given the therapy. Let X_1, X_2, and X_3 denote the number of patients who improved, stayed the same, or became degraded, respectively. Determine the following:

a. Are X_1, X_2, X_3 independent?
b. $P(X_1 = 10)$
c. $P(X_1 = 10, X_2 = 8, X_3 = 2)$
d. $P(X_1 = 5 \mid X_2 = 12)$
e. $E(X_1)$

Exercises for Section 5.6

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

5.6.1 X and Y are independent, normal random variables with $E(X) = 2, V(X) = 5, E(Y) = 6$, and $V(Y) = 8$.
Determine the following:

a. $E(3X + 2Y)$
b. $V(3X + 2Y)$
c. $P(3X + 2Y < 18)$
d. $P(3X + 2Y < 28)$

5.6.2 **WP** Suppose that the random variable X represents the length of a punched part in centimeters. Let Y be the length of the part in millimeters. If $E(X) = 5$ and $V(X) = 0.25$, what are the mean and variance of Y?

5.6.3 **WP** **SS** In the manufacture of electroluminescent lamps, several different layers of ink are deposited onto a plastic substrate. The thickness of these layers is critical if specifications regarding the final color and intensity of light are to be met. Let X and Y denote the thickness of two different layers of ink. It is known that X is normally distributed with a mean of 0.1 mm and a standard deviation of 0.00031 mm, and Y is also normally distributed with a mean of 0.23 mm and a standard deviation of 0.00017 mm. Assume that these variables are independent.

a. If a particular lamp is made up of these two inks only, what is the probability that the total ink thickness is less than 0.2337 mm?
b. A lamp with a total ink thickness exceeding 0.2405 mm lacks the uniformity of color that the customer demands. Find the probability that a randomly selected lamp fails to meet customer specifications.

5.6.4 **WP** Making handcrafted pottery generally takes two major steps: wheel throwing and firing. The time of wheel throwing and the time of firing are normally distributed random variables with means of 40 minutes and 60 minutes and standard deviations of 2 minutes and 3 minutes, respectively.

a. What is the probability that a piece of pottery will be finished within 95 minutes?
b. What is the probability that it will take longer than 110 minutes?

5.6.5 **WP** An article in *Knee Surgery Sports Traumatology, Arthroscopy* ["Effect of Provider Volume on Resource Utilization for Surgical Procedures" (2005, Vol. 13, pp. 273–279)] showed a mean time of 129 minutes and a standard deviation of 14 minutes for ACL reconstruction surgery for high-volume hospitals (with more than 300 such surgeries per year). If a high-volume hospital needs to schedule 10 surgeries, what are the mean and variance of the total time to complete these surgeries? Assume that the times of the surgeries are independent and normally distributed.

5.6.6 **WP** An automated filling machine fills soft-drink cans, and the standard deviation is 0.5 fluid ounce. Assume that the fill volumes of the cans are independent, normal random variables.

a. What is the standard deviation of the average fill volume of 100 cans?
b. If the mean fill volume is 12.1 oz, what is the probability that the average fill volume of 100 cans is less than 12 oz?
c. What should the mean fill volume equal so that the probability that the average of 100 cans is less than 12 oz is 0.005?
d. If the mean fill volume is 12.1 oz, what should the standard deviation of fill volume equal so that the probability that the average of 100 cans is less than 12 oz is 0.005?
e. Determine the number of cans that need to be measured such that the probability that the average fill volume is less than 12 oz is 0.01.

5.6.7 **WP** **SS** **VS** Assume that the weights of individuals are independent and normally distributed with a mean of 160 pounds and a standard deviation of 30 pounds. Suppose that 25 people squeeze into an elevator that is designed to hold 4300 pounds.

a. What is the probability that the load (total weight) exceeds the design limit?
b. What design limit is exceeded by 25 occupants with probability 0.0001?

5.6.8 A U-shaped component is to be formed from the three parts A, B, and C. See Figure 5.14. The length of A is normally distributed with a mean of 10 mm and a standard deviation of 0.1 mm. The thicknesses of parts B and C are each normally distributed with a mean of 2 mm and a standard deviation of 0.05 mm. Assume that all dimensions are independent.

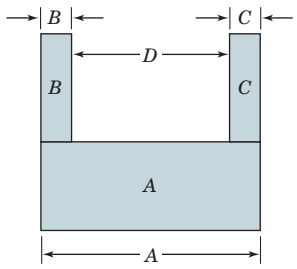


FIGURE 5.14
Illustration for the U-shaped component.

a. Determine the mean and standard deviation of gap D.
b. What is the probability that gap D is less than 5.9 mm?

5.6.9 Consider the perimeter of a part in Example 5.27. Let X_1 and X_2 denote the length and width of a part with standard deviations 0.1 and 0.2 centimeters, respectively. Suppose that the covariance between X_1 and X_2 is 0.02. Determine the variance of the perimeter $Y = 2X_1 + 2X_2$ of a part. Compare and comment on the result here and in the example.

5.6.10 The rate of return of an asset is the change in price divided by the initial price (denoted as r). Suppose that \$10,000 is used to purchase shares in three stocks with rates of returns X_1, X_2, X_3. Initially, \$2500, \$3000, and \$4500 are allocated to each one, respectively. After one year, the distribution of the rate of return for each is normally distributed with the following parameters: $\mu_1 = 0.12$, $\sigma_1 = 0.14$, $\mu_2 = 0.04$, $\sigma_2 = 0.02$, $\mu_3 = 0.07$, $\sigma_3 = 0.08$.

a. Assume that these rates of return are independent. Determine the mean and variance of the rate of return after one year for the entire investment of \$10,000.
b. Assume that X_1 is independent of X_2 and X_3 but that the covariance between X_2 and X_3 is -0.005. Repeat part (a).
c. Compare the means and variances obtained in parts (a) and (b) and comment on any benefits from negative covariances between the assets.

5.6.11 In Exercise 5.2.7, the monthly demand for MMR vaccine was assumed to be approximately normally distributed with a mean and standard deviation of 1.1 and 0.3 million doses, respectively. Suppose that the demands for different months are independent, and let Z denote the demand for a year (in millions of doses). Determine the following:

a. Mean, variance, and distribution of Z
b. $P(Z < 13.2)$
c. $P(11 < Z < 15)$
d. Value for c such that $P(Z < c) = 0.99$

Exercises for Section 5.7

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

5.7.1 WP Suppose that X is a random variable with probability distribution

$$f_X(x) = 1/4, \qquad x = 1, 2, 3, 4$$

Determine the probability distribution of $Y = 2X + 1$.

5.7.2 WP Suppose that X is a continuous random variable with probability distribution

$$f_X(x) = \frac{x}{18}, \qquad 0 \le x \le 6$$

a. Determine the probability distribution of the random variable $Y = 2X + 10$.
b. Determine the expected value of Y.

5.7.3 WP A random variable X has the probability distribution

$$f_X(x) = e^{-x}, \qquad x \ge 0$$

Determine the probability distribution for the following:

a. $Y = X^2$ **b.** $Y = X^{1/2}$ **c.** $Y = \ln X$

5.7.4 WP The velocity of a particle in a gas is a random variable V with probability distribution

$$f_V(v) = av^2e^{-bv}, \qquad v > 0$$

where b is a constant that depends on the temperature of the gas and the mass of the particle.

a. Determine the value of the constant a.
b. The kinetic energy of the particle is $W = mV^2/2$. Determine the probability distribution of W.

5.7.5 WP Suppose that X has the probability distribution

$$f_X(x) = 1, \qquad 1 \le x \le 2$$

Determine the probability distribution of the random variable $Y = e^X$.

5.7.6 Derive the probability density function for a lognormal random variable Y from the relationship that $Y = \exp(W)$ for a normal random variable W with mean θ and variance ω^2.

5.7.7 SS An aircraft is flying at a constant altitude with velocity magnitude r_1 (relative to the air) and angle θ_1 (in a two-dimensional coordinate system). The magnitude and direction of the wind are r_2 and θ_2, respectively. Suppose that the wind angle is uniformly distributed between 10 and 20 degrees and all other parameters are constant. Determine the probability density function of the magnitude of the resultant vector $r = [r_1^2 + r_2^2 + r_1r_2 (\cos\theta_1 - \cos\theta_2)]^{0.5}$.

5.7.8 Power meters enable cyclists to obtain power measurements nearly continuously. The meters also calculate the average power generated over a time interval. Professional riders can generate 6.6 watts per kilogram of body weight for extended periods of time. Some meters calculate a normalized power measurement to

adjust for the physiological effort required when the power output changes frequently. Let the random variable X denote the power output at a measurement time and assume that X has a lognormal distribution with parameters $\theta = 5.2933$ and $\omega^2 = 0.00995$. The normalized power is computed as the fourth root of the mean of $Y = X^4$. Determine the following:

a. Mean and standard deviation of X
b. $f_Y(y)$
c. Mean and variance of Y
d. Fourth root of the mean of Y
e. Compare $[E(X^4)]^{1/4}$ to $E(X)$ and comment.

Exercises for Section 5.8

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

5.8.1 A random variable X has the discrete uniform distribution

$$f(x) = \frac{1}{m}, \quad x = 1, 2, \ldots, m$$

a. Show that the moment-generating function is

$$M_X(t) = \frac{e^t(1 - e^{tm})}{m(1 - e^t)}$$

b. Use $M_X(t)$ to find the mean and variance of X.

5.8.2 A random variable X has the Poisson distribution

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots$$

a. Show that the moment-generating function is

$$M_X(t) = e^{\lambda(e^t - 1)}$$

b. Use $M_X(t)$ to find the mean and variance of the Poisson random variable.

5.8.3 The geometric random variable X has probability distribution

$$f(x) = (1 - p)^{x-1}p, \quad x = 1, 2, \ldots$$

a. Show that the moment-generating function is

$$M_X(t) = \frac{pe^t}{1 - (1 - p)e^t}$$

b. Use $M_X(t)$ to find the mean and variance of X.

5.8.4 The chi-squared random variable with k degrees of freedom has moment-generating function $M_X(t) = (1 - 2t)^{-k/2}$. Suppose that X_1 and X_2 are independent chi-squared random variables with k_1 and k_2 degrees of freedom, respectively. What is the distribution of $Y = X_1 + X_2$?

5.8.5 **SS** A continuous random variable X has the following probability distribution:

$$f(x) = 4xe^{-2x}, \quad x > 0$$

a. Find the moment-generating function for X.
b. Find the mean and variance of X.

5.8.6 The continuous uniform random variable X has density function

$$f(x) = \frac{1}{\beta - \alpha}, \quad \alpha \le x \le \beta$$

a. Show that the moment-generating function is

$$M_X(t) = \frac{e^{t\beta} - e^{t\alpha}}{t(\beta - \alpha)}$$

b. Use $M_X(t)$ to find the mean and variance of X.

5.8.7 A random variable X has the exponential distribution

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

a. Show that the moment-generating function for $t < \lambda$ of X is

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)$$

b. Find the mean and variance of X.

5.8.8 A random variable X has the gamma distribution

$$f(x) = \frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1}e^{-\lambda x}, \quad x > 0$$

a. Show that the moment-generating function for $t < \lambda$ of X is

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-r}$$

b. Find the mean and variance of X.

5.8.9 Let $X_1, X_2, \ldots, X_r$ be independent exponential random variables with parameter λ.

a. Find the moment-generating function of $Y = X_1 + X_2 + \ldots + X_r$.
b. What is the distribution of the random variable Y?

5.8.10 Suppose that X_i has a normal distribution with mean μ_i and variance σ_i^2, $i = 1, 2$. Let X_1 and X_2 be independent.

a. Find the moment-generating function of $Y = X_1 + X_2$.
b. What is the distribution of the random variable Y?

Supplemental Exercises for Chapter 5

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

5.S11 WP Show that the following function satisfies the properties of a joint probability mass function:

x	y	$f(x, y)$
0	0	1/4
0	1	1/8
1	0	1/8
1	1	1/4
2	2	1/4

Determine the following:

a. $P(X < 0.5, Y < 1.5)$ **b.** $P(X \leq 1)$
c. $P(X < 1.5)$ **d.** $P(X > 0.5, Y < 1.5)$
e. $E(X), E(Y), V(X), V(Y)$
f. Marginal probability distribution of the random variable X
g. Conditional probability distribution of Y given that $X = 1$
h. $E(Y \mid X = 1)$
i. Are X and Y independent? Why or why not?
j. Correlation between X and Y

5.S12 WP The percentage of people given an antirheumatoid medication who suffer severe, moderate, or minor side effects are 10, 20, and 70%, respectively. Assume that people react independently and that 20 people are given the medication.
Determine the following:

a. Probability that 2, 4, and 14 people will suffer severe, moderate, or minor side effects, respectively
b. Probability that no one will suffer severe side effects
c. Mean and variance of the number of people who will suffer severe side effects
d. Conditional probability distribution of the number of people who suffer severe side effects given that 19 suffer minor side effects
e. Conditional mean of the number of people who suffer severe side effects given that 19 suffer minor side effects

5.S13 WP SS Determine the value of c such that the function $f(x, y) = cx^2y$ for $0 < x < 3$ and $0 < y < 2$ satisfies the properties of a joint probability density function.
Determine the following:

a. $P(X < 1, Y < 1)$ **b.** $P(X < 2.5)$
c. $P(1 < Y < 2.5)$ **d.** $P(X > 2.1 < Y < 1.5)$
e. $E(X)$ **f.** $E(Y)$
g. Marginal probability distribution of the random variable X
h. Conditional probability distribution of Y given that $X = 1$
i. Conditional probability distribution of X given that $Y = 1$

5.S14 The joint distribution of the continuous random variables X, Y, and Z is constant over the region $x^2 + y^2 \leq 1$, $0 < z < 4$. Determine the following:

a. $P(X^2 + Y^2 \leq 0.5)$ **b.** $P(X^2 + Y^2 \leq 0.5, Z < 2)$
c. Joint conditional probability density function of X and Y given that $Z = 1$
d. Marginal probability density function of X
e. Conditional mean of Z given that $X = 0$ and $Y = 0$
f. Conditional mean of Z given that $X = x$ and $Y = y$

5.S15 WP Contamination problems in semiconductor manufacturing can result in a functional defect, a minor defect, or no defect in the final product. Suppose that 20%, 50%, and 30% of the contamination problems result in functional, minor, and no defects, respectively. Assume that the defects of 10 contamination problems are independent.

a. What is the probability that the 10 contamination problems result in two functional defects and five minor defects?
b. What is the distribution of the number of contamination problems that result in no defects?
c. What is the expected number of contamination problems that result in no defects?

5.S16 WP The weight of adobe bricks for construction is normally distributed with a mean of 3 pounds and a standard deviation of 0.25 pound. Assume that the weights of the bricks are independent and that a random sample of 25 bricks is chosen.

a. What is the probability that the mean weight of the sample is less than 2.95 pounds?
b. What value will the mean weight exceed with probability 0.99?

5.S17 WP The time for an automated system in a warehouse to locate a part is normally distributed with a mean of 45 seconds and a standard deviation of 30 seconds. Suppose that independent requests are made for 10 parts.

a. What is the probability that the average time to locate 10 parts exceeds 60 seconds?
b. What is the probability that the total time to locate 10 parts exceeds 600 seconds?

5.S18 WP The weight of a small candy is normally distributed with a mean of 0.1 ounce and a standard deviation of 0.01 ounce. Suppose that 16 candies are placed in a package and that the weights are independent.

a. What are the mean and variance of the package's net weight?
b. What is the probability that the net weight of a package is less than 1.6 ounces?
c. What value will the mean weight exceed with probability 0.99?

5.S19 Suppose that X and Y have a bivariate normal distribution with $\sigma_X = 4$, $\sigma_Y = 1$, $\mu_X = 2$, $\mu_Y = 4$, and $\rho = -0.2$. Draw a rough contour plot of the joint probability density function.

5.S20 WP A mechanical assembly used in an automobile engine contains four major components. The weights of the

components are independent and normally distributed with the following means and standard deviations (in ounces):

Component	Mean	Standard Deviation
Left case	4.0	0.4
Right case	5.5	0.5
Bearing assembly	10.0	0.2
Bolt assembly	8.0	0.5

a. What is the probability that the weight of an assembly exceeds 29.5 ounces?

b. What is the probability that the mean weight of eight independent assemblies exceeds 29 ounces?

5.S21 WP The permeability of a membrane used as a moisture barrier in a biological application depends on the thickness of two integrated layers. The layers are normally distributed with means of 0.5 and 1 millimeters, respectively. The standard deviations of layer thickness are 0.1 and 0.2 millimeters, respectively. The correlation between layers is 0.7.

a. Determine the mean and variance of the total thickness of the two layers.

b. What is the probability that the total thickness is less than 1 millimeter?

c. Let X_1 and X_2 denote the thickness of layers 1 and 2, respectively. A measure of performance of the membrane is a function of $2X_1 + 3X_2$ of the thickness. Determine the mean and variance of this performance measure.

5.S22 A small company is to decide what investments to use for cash generated from operations. Each investment has a mean and standard deviation associated with a percentage return. The first security has a mean return of 5% with a standard deviation of 2%, and the second security provides the same mean of 5% with a standard deviation of 4%. The securities have a correlation of –0.5. If the company invests $2 million with half in each security, what are the mean and standard deviation of the percentage return? Compare the standard deviation of this strategy to one that invests the $2 million into the first security only.

5.S23 WP A marketing company performed a risk analysis for a manufacturer of synthetic fibers and concluded that new competitors present no risk 13% of the time (due mostly to the diversity of fibers manufactured), moderate risk 72% of the time (some overlapping of products), and very high risk (competitor manufactures the exact same products) 15% of the time. It is known that 12 international companies are planning to open new facilities for the manufacture of synthetic fibers within the next 3 years. Assume that the companies are independent. Let X, Y, and Z denote the number of new competitors that will pose no, moderate, and very high risk for the interested company, respectively.

Determine the following:

a. Range of the joint probability distribution of X, Y, and Z

b. $P(X = 1, Y = 3, Z = 1)$ **c.** $P(Z \leq 2)$

d. $P(Z = 2 \mid Y = 1, X = 10)$ **e.** $P(Z \leq 1 \mid X = 10)$

f. $P(Y \leq 1, Z \leq 1 \mid X = 10)$ **g.** $E(Z \mid X = 10)$

5.S24 Suppose X has a lognormal distribution with parameters θ and ω. Determine the probability density function and the parameters values for $Y = X^{\gamma}$ for a constant $\gamma > 0$. What is the name of this distribution?

5.S25 The power in a DC circuit is $P = I^2/R$ where I and R denote the current and resistance, respectively. Suppose that I is approximately normally distributed with mean of 200 mA and standard deviation 0.2 mA and R is a constant. Determine the probability density function of power.

5.S26 The intensity (mW/mm^2) of a laser beam on a surface theoretically follows a bivariate normal distribution with maximum intensity at the center, equal variance σ in the x and y directions, and zero covariance. There are several definitions for the width of the beam. One definition is the diameter at which the intensity is 50% of its peak. Suppose that the beam width is 1.6 mm under this definition. Determine σ. Also determine the beam width when it is defined as the diameter where the intensity equals $1/e^2$ of the peak.

Chapter 6 Exercises

Exercises for Section 6.1

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

6.1.1 Will the sample mean always correspond to one of the observations in the sample?

6.1.2 Will exactly half of the observations in a sample fall below the mean?

6.1.3 Will the sample mean always be the most frequently occurring data value in the sample?

6.1.4 For any set of data values, is it possible for the sample standard deviation to be larger than the sample mean? If so, give an example.

6.1.5 Can the sample standard deviation be equal to zero? If so, give an example.

6.1.6 WP GO Tutorial In *Applied Life Data Analysis* (Wiley, 1982), Wayne Nelson presents the breakdown time of an insulating fluid between electrodes at 34 kV. The times, in minutes, are as

follows: 0.19, 0.78, 0.96, 1.31, 2.78, 3.16, 4.15, 4.67, 4.85, 6.50, 7.35, 8.01, 8.27, 12.06, 31.75, 32.52, 33.91, 36.71, and 72.89. Calculate the sample mean and sample standard deviation.

6.1.7 WP SS VS An article in *Human Factors* ["Visual Performance on CRT Screens and Hard-Copy Displays" (1989, Vol. 31(3), pp. 247–257)] presented data on visual accommodation (a function of eye movement) when recognizing a speckle pattern on a high-resolution CRT screen. The data are as follows: 36.45, 67.90, 38.77, 42.18, 26.72, 50.77, 39.30, and 49.71. Calculate the sample mean and sample standard deviation. Construct a dot diagram of the data.

6.1.8 Preventing fatigue crack propagation in aircraft structures is an important element of aircraft safety. An engineering study to investigate fatigue crack in $n = 9$ cyclically loaded wing boxes reported the following crack lengths (in mm): 2.13, 2.96, 3.02, 1.82, 1.15, 1.37, 2.04, 2.47, 2.60. Calculate the sample mean and sample standard deviation. Prepare a dot diagram of the data.

6.1.9 An article in the *Journal of Physiology* ["Response of Rat Muscle to Acute Resistance Exercise Defined by Transcriptional and Translational Profiling" (2002, Vol. 545, pp. 27–41)] studied gene expression as a function of resistance exercise. Expression data (measures of gene activity) from one gene are shown in the following table. One group of rats was exercised for 6 hours while the other received no exercise. Compute the sample mean and standard deviation of the exercise and no-exercise groups separately. Construct a dot diagram for the exercise and no-exercise groups separately. Comment on any differences for the groups.

6 Hours of Exercise	6 Hours of Exercise	No Exercise	No Exercise
425.313	208.475	485.396	406.921
223.306	286.484	159.471	335.209
388.793	244.242	478.314	
139.262	408.099	245.782	
212.565	157.743	236.212	
324.024	436.37	252.773	

6.1.10 WP Exercise 6.1.7 describes data from an article in *Human Factors* on visual accommodation from an experiment involving a high-resolution CRT screen.

Data from a second experiment using a low-resolution screen were also reported in the article. They are 8.85, 35.80, 26.53, 64.63, 9.00, 15.38, 8.14, and 8.24. Prepare a dot diagram for this second sample and compare it to the one for the first sample. What can you conclude about CRT resolution in this situation?

6.1.11 WP SS The following data are the joint temperatures of the O-rings (°F) for each test firing or actual launch of the Space Shuttle rocket motor (from *Presidential Commission on the Space Shuttle Challenger Accident*, Vol. 1, pp. 129–131): 84, 49, 61, 40, 83, 67, 45, 66, 70, 69, 80, 58, 68, 60, 67, 72, 73, 70, 57, 63, 70, 78, 52, 67, 53, 67, 75, 61, 70, 81, 76, 79, 75, 76, 58, 31.

a. Compute the sample mean and sample standard deviation and construct a dot diagram of the temperature data.

b. Set aside the smallest observation (31°F) and recompute the quantities in part (a). Comment on your findings. How "different" are the other temperatures from this last value?

6.1.12 Cloud seeding, a process in which chemicals such as silver iodide and frozen carbon dioxide are introduced by aircraft into clouds to promote rainfall, was widely used in the 20th century. Recent research has questioned its effectiveness ["Reassessment of Rain Enhancement Experiments and Operations in Israel Including Synoptic Considerations," *Journal of Atmospheric Research* (2010, Vol. 97(4), pp. 513–525)]. An experiment was performed by randomly assigning 52 clouds to be seeded or not. The amount of rain generated was then measured in acre-feet. Here are the data for the unseeded and seeded clouds:

Unseeded:

81.2 26.1 95.0 41.1 28.6 21.7 11.5 68.5 345.5 321.2
1202.6 1.0 4.9 163.0 372.4 244.3 47.3 87.0 26.3 24.4
830.1 4.9 36.6 147.8 17.3 29.0

Seeded:

274.7 302.8 242.5 255.0 17.5 115.3 31.4 703.4 334.1
1697.8 118.3 198.6 129.6 274.7 119.0 1656.0 7.7 430.0
40.6 92.4 200.7 32.7 4.1 978.0 489.1 2745.6

Find the sample mean, sample standard deviation, and range of rainfall for

a. All 52 clouds
b. The unseeded clouds
c. The seeded clouds

6.1.13 Construct dot diagrams of the seeded and unseeded clouds and compare their distributions in a couple of sentences.

6.1.14 In the 2000 Sydney Olympics, a special program initiated by the IOC president allowed developing countries to send athletes to the Olympics without the usual qualifying procedure. Here are the 71 times for the first round of the 100-meter men's swim (in seconds).

60.39 49.93 53.40 51.82 50.46 51.34 50.28 50.19 52.14
50.56 52.72 50.95 49.74 49.16 52.57 52.53 52.09 52.40
49.75 54.06 53.50 50.63 51.93 51.62 52.58 53.55 51.07
49.76 49.73 50.90 59.26 49.29 52.78 112.72 49.79 49.83
52.43 51.28 52.22 49.76 49.70 52.90 50.19 54.33 62.45
51.93 52.24 52.82 50.96 48.64 51.11 50.87 52.18 54.12
50.49 49.84 52.91 52.52 50.32 51.52 52.0 52.85 52.24
49.45 51.28 49.09 58.79 49.74 49.32 50.62 49.45

a. Find the sample mean and sample standard deviation of these 100-meter swim times.
b. Construct a dot diagram of the data.
c. Comment on anything unusual that you see.

Exercises for Section 6.2

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

6.2.1 A back-to-back stem-and-leaf display on two data sets is conducted by hanging the data on both sides of the same stems. Here is a back-to-back stem-and-leaf display for the cloud seeding data in Exercise 6.1.12 showing the unseeded clouds on the left and the seeded clouds on the right.

65098754433332221000	0	01233492223
	2	00467703
	4	39
	6	0
3	8	8
	10	
0	12	
	14	
	16	60
	18	
	20	
	22	
	24	
	26	5

How does the back-to-back stem-and-leaf display show the differences in the data set in a way that the dotplot cannot?

6.2.2 When will the median of a sample be equal to the sample mean?

6.2.3 When will the median of a sample be equal to the mode?

6.2.4 **WP** An article in *Technometrics* ["Validation of Regression Models: Methods and Examples" (1977, Vol. 19(4), p. 425)] presented the following data on the motor fuel octane ratings of several blends of gasoline:

88.5	98.8	89.6	92.2	92.7	88.4	87.5	90.9
94.7	88.3	90.4	83.4	87.9	92.6	87.8	89.9
84.3	90.4	91.6	91.0	93.0	93.7	88.3	91.8
90.1	91.2	90.7	88.2	94.4	96.5	89.2	89.7
89.0	90.6	88.6	88.5	90.4	84.3	92.3	92.2
89.8	92.2	88.3	93.3	91.2	93.2	88.9	
91.6	87.7	94.2	87.4	86.7	88.6	89.8	
90.3	91.1	85.3	91.1	94.2	88.7	92.7	
90.0	86.7	90.1	90.5	90.8	92.7	93.3	
91.5	93.4	89.3	100.3	90.1	89.3	86.7	
89.9	96.1	91.1	87.6	91.8	91.0	91.0	

Construct a stem-and-leaf display for these data. Calculate the median and quartiles of these data.

6.2.5 **WP GO Tutorial** **SS** The following data are the numbers of cycles to failure of aluminum test coupons subjected to repeated alternating stress at 21,000 psi, 18 cycles per second.

1115	865	1015	885	1594	1000	1416	1501
1310	2130	845	1223	2023	1820	1560	1238
1540	1421	1674	375	1315	1940	1055	990
1502	1109	1016	2265	1269	1120	1764	1468
1258	1481	1102	1910	1260	910	1330	1512
1315	1567	1605	1018	1888	1730	1608	1750
1085	1883	706	1452	1782	1102	1535	1642
798	1203	2215	1890	1522	1578	1781	
1020	1270	785	2100	1792	758	1750	

Construct a stem-and-leaf display for these data. Calculate the median and quartiles of these data. Does it appear likely that a coupon will "survive" beyond 2000 cycles? Justify your answer.

6.2.6 **WP** Calculate the sample median, mode, and mean of the data in Exercise 6.2.4. Explain how these three measures of location describe different features of the data.

6.2.7 Calculate the sample median, mode, and mean of the data in Exercise 6.2.5. Explain how these three measures of location describe different features in the data.

6.2.8 **WP** The female students in an undergraduate engineering core course at ASU self-reported their heights to the nearest inch. The data follow. Construct a stem-and-leaf diagram for the height data and comment on any important features that you notice. Calculate the sample mean, the sample standard deviation, and the sample median of height.

62	64	61	67	65	68	61	65	60	65	64	63	59
68	64	66	68	69	65	67	62	66	68	67	66	65
69	65	69	65	67	67	65	63	64	67	65		

6.2.9 The shear strengths of 100 spot welds in a titanium alloy follow. Construct a stem-and-leaf diagram for the weld strength data and comment on any important features that you notice. What is the 95th percentile of strength?

5408	5431	5475	5442	5376	5388	5459	5422	5416	5435
5420	5429	5401	5446	5487	5416	5382	5357	5388	5457
5407	5469	5416	5377	5454	5375	5409	5459	5445	5429
5463	5408	5481	5453	5422	5354	5421	5406	5444	5466
5399	5391	5477	5447	5329	5473	5423	5441	5412	5384
5445	5436	5454	5453	5428	5418	5465	5427	5421	5396
5381	5425	5388	5388	5378	5481	5387	5440	5482	5406
5401	5411	5399	5431	5440	5413	5406	5342	5452	5420
5458	5485	5431	5416	5431	5390	5399	5435	5387	5462
5383	5401	5407	5385	5440	5422	5448	5366	5430	5418

6.2.10 An important quality characteristic of water is the concentration of suspended solid material. Following are

60 measurements on suspended solids from a certain lake. Construct a stem-and-leaf diagram for these data and comment on any important features that you notice. Compute the sample mean, the sample standard deviation, and the sample median. What is the 90th percentile of concentration?

42.4	65.7	29.8	58.7	52.1	55.8	57.0	68.7	67.3	67.3
54.3	54.0	73.1	81.3	59.9	56.9	62.2	69.9	66.9	59.0
56.3	43.3	57.4	45.3	80.1	49.7	42.8	42.4	59.6	65.8
61.4	64.0	64.2	72.6	72.5	46.1	53.1	56.1	67.2	70.7
42.6	77.4	54.7	57.1	77.3	39.3	76.4	59.3	51.1	73.8
61.4	73.1	77.3	48.5	89.8	50.7	52.0	59.6	66.1	31.6

6.2.11 WP A group of wine enthusiasts taste-tested a pinot noir wine from Oregon. The evaluation was to grade the wine on a 0-to-100-point scale. The results follow. Construct a stem-and-leaf diagram for these data and comment on any important features that you notice. Compute the sample mean, the sample standard deviation, and the sample median. A wine rated above 90 is considered truly exceptional. What proportion of the taste-tasters considered this particular pinot noir truly exceptional?

94	90	92	91	91	86	89	91	91	90
90	93	87	90	91	92	89	86	89	90
88	95	91	88	89	92	87	89	95	92
85	91	85	89	88	84	85	90	90	83

6.2.12 WP In their book *Introduction to Linear Regression Analysis* (5th edition, Wiley, 2012), Montgomery, Peck, and Vining presented measurements on $NbOCl_3$ concentration from a tube-flow reactor experiment. The data, in gram-mole per liter $\times 10^{-3}$, are as follows. Construct a stem-and-leaf diagram for these data and comment on any important features that you notice. Compute the sample mean, the sample standard deviation, and the sample median.

450	450	473	507	457	452	453	1215	1256
1145	1085	1066	1111	1364	1254	1396	1575	1617
1733	2753	3186	3227	3469	1911	2588	2635	2725

6.2.13 In Exercise 6.2.8, we presented height data that were self-reported by female undergraduate engineering students in a core course at ASU. In the same class, the male students self-reported their heights as follows. Construct a comparative stem-and-leaf diagram by listing the stems in the center of the display and then placing the female leaves on the left and the male leaves on the right. Comment on any important features that you notice in this display.

69	67	69	70	65	68	69	70	71	69	66	67	69	75	68	67	68
69	70	71	72	68	69	69	70	71	68	72	69	69	68	69	73	70
73	68	69	71	67	68	65	68	68	69	70	74	71	69	70	69	

Exercises for Section 6.3

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

6.3.1 VS Construct a frequency distribution and histogram using the failure data from Exercise 6.2.5.
6.3.2 WP Construct a frequency distribution and histogram for the motor fuel octane data from Exercise 6.2.4. Use eight bins.
6.3.3 SS Construct histograms with 8 and 16 bins for the data in Exercise 6.2.5. Compare the histograms. Do both histograms display similar information?
6.3.4 WP VS Construct frequency distributions and histograms with 8 bins and 16 bins for the motor fuel octane data in Exercise 6.2.4. Compare the histograms. Do both histograms display similar information?
6.3.5 Construct a histogram for the spot weld shear strength data in Exercise 6.2.9. Comment on the shape of the histogram. Does it convey the same information as the stem-and-leaf display?
6.3.6 WP Construct a histogram for the female student height data in Exercise 6.2.8.
6.3.7 Construct a histogram for the pinot noir wine rating data in Exercise 6.2.11. Comment on the shape of the histogram. Does it convey the same information as the stem-and-leaf display?
6.3.8 WP Construct a histogram for the water quality data in Exercise 6.2.10. Comment on the shape of the histogram. Does it convey the same information as the stem-and-leaf display?
6.3.9 WP SS VS **The Pareto Chart**. An important variation of a histogram for categorical data is the Pareto chart. This chart is widely used in quality improvement efforts, and the categories usually represent different types of defects, failure modes, or product/process problems. The categories are ordered so that the category with the largest frequency is on the left, followed by the category with the second largest frequency, and so forth. These charts are named after Italian economist Vilfredo Pareto, and they usually exhibit "Pareto's law"; that is, most of the defects can be accounted for by only a few categories. Suppose that the following information on structural defects in automobile doors is obtained: dents, 4; pits, 4; parts assembled out of sequence, 6; parts undertrimmed, 21; missing holes/slots, 8; parts not lubricated, 5; parts out of contour, 30; and parts not deburred, 3. Construct and interpret a Pareto chart.

Exercises for Section 6.4

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

6.4.1 Using the data from Exercise 6.1.12 on cloud seeding,
- **a.** Find the median and quartiles for the unseeded cloud data.
- **b.** Find the median and quartiles for the seeded cloud data.
- **c.** Make two side-by-side box plots, one for each group on the same plot.
- **d.** Compare the distributions from what you can see in the side-by-side box plots.

6.4.2 Using the data from Exercise 6.1.14 on swim times,
- **a.** Find the median and quartiles for the data.
- **b.** Make a box plot of the data.
- **c.** Repeat (a) and (b) for the data without the extreme outlier and comment.
- **d.** Compare the distribution of the data with and without the extreme outlier.

6.4.3 WP GO Tutorial SS VS The "cold start ignition time" of an automobile engine is being investigated by a gasoline manufacturer. The following times (in seconds) were obtained for a test vehicle: 1.75, 1.92, 2.62, 2.35, 3.09, 3.15, 2.53, 1.91.
- **a.** Calculate the sample mean, sample variance, and sample standard deviation.
- **b.** Construct a box plot of the data.

6.4.4 An article in *Transactions of the Institution of Chemical Engineers* (1956, Vol. 34, pp. 280–293) reported data from an experiment investigating the effect of several process variables on the vapor phase oxidation of naphthalene. A sample of the percentage mole conversion of naphthalene to maleic anhydride follows: 4.2, 4.7, 4.7, 5.0, 3.8, 3.6, 3.0, 5.1, 3.1, 3.8, 4.8, 4.0, 5.2, 4.3, 2.8, 2.0, 2.8, 3.3, 4.8, 5.0.
- **a.** Calculate the sample mean, sample variance, and sample standard deviation.
- **b.** Construct a box plot of the data.

6.4.5 Exercise 6.1.11 presented the joint temperatures of the O-rings (°F) for each test firing or actual launch of the Space Shuttle rocket motor. In that exercise, you were asked to find the sample mean and sample standard deviation of temperature.
- **a.** Find the median and the upper and lower quartiles of temperature.
- **b.** Set aside the lowest observation (31°F) and recompute the quantities in part (a). Comment on your findings. How "different" are the other temperatures from this lowest value?
- **c.** Construct a box plot of the data and comment on the possible presence of outliers.

6.4.6 WP Reconsider the water quality data in Exercise 6.2.10. Construct a box plot of the concentrations and write an interpretation of the plot. How does the box plot compare in interpretive value to the original stem-and-leaf diagram?

6.4.7 Reconsider the weld strength data in Exercise 6.2.9. Construct a box plot of the data and write an interpretation of the plot. How does the box plot compare in interpretive value to the original stem-and-leaf diagram?

6.4.8 WP In Exercise 6.4.3, data were presented on the cold start ignition time of a particular gasoline used in a test vehicle. A second formulation of the gasoline was tested in the same vehicle with the following times (in seconds): 1.83, 1.99, 3.13, 3.29, 2.65, 2.87, 3.40, 2.46, 1.89, and 3.35. Use these new data along with the cold start times reported in Exercise 6.4.3 to construct comparative box plots. Write an interpretation of the information that you see in these plots.

6.4.9 An article in *Nature Genetics* ["Treatment-specific Changes in Gene Expression Discriminate In Vivo Drug Response in Human Leukemia Cells" (2003, Vol. 34(1), pp. 85–90)] studied gene expression as a function of treatments for leukemia. One group received a high dose of the drug, while the control group received no treatment. Expression data (measures of gene activity) from one gene are shown in the table on the next page. Construct a box plot for each group of patients. Write an interpretation to compare the information in these plots.

High Dose	Control	Control	Control
16.1	297.1	25.1	131.1
134.9	491.8	820.1	166.5
52.7	1332.9	82.5	2258.4
14.4	1172	713.9	497.5
124.3	1482.7	785.6	263.4
99	335.4	114	252.3
24.3	528.9	31.9	351.4
16.3	24.1	86.3	678.9
15.2	545.2	646.6	3010.2
47.7	92.9	169.9	67.1
12.9	337.1	20.2	318.2
72.7	102.3	280.2	2476.4
126.7	255.1	194.2	181.4
46.4	100.5	408.4	2081.5
60.3	159.9	155.5	424.3
23.5	168	864.6	188.1
43.6	95.2	355.4	563
79.4	132.5	634	149.1
38	442.6	2029.9	2122.9
58.2	15.8	362.1	1295.9
26.5	175.6		

Exercises for Section 6.5

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

6.5.1 **WP** In their book *Time Series Analysis, Forecasting, and Control* (Prentice Hall, 1994), G. E. P. Box, G. M. Jenkins, and G. C. Reinsel present chemical concentration readings made every 2 hours. Some of these data follow (read down the column, then left to right).

17.0	16.7	17.1	17.5	17.6
16.6	17.4	17.4	18.1	17.5
16.3	17.2	17.4	17.5	16.5
16.1	17.4	17.5	17.4	17.8
17.1	17.4	17.4	17.4	17.3
16.9	17.0	17.6	17.1	17.3
16.8	17.3	17.4	17.6	17.1
17.4	17.2	17.3	17.7	17.4
17.1	17.4	17.0	17.4	16.9
17.0	16.8	17.8	17.8	17.3

Construct and interpret either a digidot plot or a separate stem-and-leaf and time series plot of these data.

6.5.2 The following table shows U.S. petroleum imports as a percentage of the totals, and Persian Gulf imports as a percentage of all imports by year since 1973 (*Source*: U.S. Department of Energy Web site, *www.eia.doe.gov*). Construct and interpret either a digidot plot or a separate stem-and-leaf and time series plot for each column of data.

Year	Petroleum Imports (thousand barrels per day)	Total Petroleum Imports as Percent of Petroleum Products Supplied	Petroleum Imports from Persian Gulf as Percent of Total Petroleum Imports
1973	6256	36.1	13.5
1974	6112	36.7	17.0
1975	6055	37.1	19.2
1976	7313	41.8	25.1
1977	8807	47.7	27.8
1978	8363	44.3	26.5
1979	8456	45.6	24.4
1980	6909	40.5	21.9
1981	5996	37.3	20.3
1982	5113	33.4	13.6
1983	5051	33.1	8.7
1984	5437	34.5	9.3
1985	5067	32.2	6.1
1986	6224	38.2	14.6
1987	6678	40.0	16.1
1988	7402	42.8	20.8
1989	8061	46.5	23.0
1990	8018	47.1	24.5
1991	7627	45.6	24.1
1992	7888	46.3	22.5
1993	8620	50.0	20.6
1994	8996	50.7	19.2

Year	Petroleum Imports (thousand barrels per day)	Total Petroleum Imports as Percent of Petroleum Products Supplied	Petroleum Imports from Persian Gulf as Percent of Total Petroleum Imports
1995	8835	49.8	17.8
1996	9478	51.7	16.9
1997	10,162	54.5	17.2
1998	10,708	56.6	19.9
1999	10,852	55.5	22.7
2000	11,459	58.1	21.7
2001	11,871	60.4	23.2
2002	11,530	58.3	19.6
2003	12,264	61.2	20.3
2004	13,145	63.4	18.9
2005	13,714	65.9	17.0
2006	13,707	66.3	16.1
2007	13,468	65.1	16.1
2008	12,915	66.2	18.4
2009	11,691	62.3	14.4
2010	11,793	61.5	14.5
2011	11,436	60.6	16.3
2012	10,627	57.3	20.3
2013	9859	52.0	20.4
2014	9241	48.4	20.3
2015	9449	48.4	15.9

6.5.3 WP SS The 100 annual Wolfer sunspot numbers from 1770 to 1869 follow. (For an interesting analysis and interpretation of these numbers, see the book by Box, Jenkins, and Reinsel referenced in Exercise 6.5.1. Their analysis requires some advanced knowledge of statistics and statistical model building.) Read down, then left to right. The 1869 result is 74. Construct and interpret either a digidot plot or a stem-and-leaf and time series plot of these data.

101	31	154	38	83	90
82	7	125	23	132	67
66	20	85	10	131	60
35	92	68	24	118	47
41	10	16	8	62	94
21	8	7	13	98	96
16	2	4	57	124	77
6	0	2	122	96	59
4	1	8	138	66	44
7	5	17	103	64	47
14	12	36	86	54	30
34	14	50	63	39	16
45	35	62	37	21	7
43	46	67	24	7	37
48	41	71	11	4	74
42	30	48	15	23	
28	24	28	40	55	

6.5.4 The table on the next page contains the global mean surface air temperature anomaly and the global CO_2 concentration for the years 1880–2004 (*Source:* http://data.giss.nasa.gov/gistemp/). The temperature is measured at a number of locations around the world and averaged annually, and then subtracted from a base period average (1951–1980) and the result reported as an anomaly.

a. Construct a time series plot of the global mean surface air temperature anomaly data and comment on any features that you observe.

b. Construct a time series plot of the global CO_2 concentration data and comment on any features that you observe.

c. Overlay the two plots on the same set of axes and comment on the plot.

Year	Anomaly, °C	CO_2, ppmv	Year	Anomaly, °C	CO_2, ppmv	Year	Anomaly, °C	CO_2, ppmv
1880	−0.11	290.7	1922	−0.09	303.8	1964	−0.25	319.2
1881	−0.13	291.2	1923	−0.16	304.1	1965	−0.15	320.0
1882	−0.01	291.7	1924	−0.11	304.5	1966	−0.07	321.1
1883	−0.04	292.1	1925	−0.15	305.0	1967	−0.02	322.0
1884	−0.42	292.6	1926	0.04	305.4	1968	−0.09	322.9
1885	−0.23	293.0	1927	−0.05	305.8	1969	0.00	324.2
1886	−0.25	293.3	1928	0.01	306.3	1970	0.04	325.2
1887	−0.45	293.6	1929	−0.22	306.8	1971	−0.10	326.1
1888	−0.23	293.8	1930	−0.03	307.2	1972	−0.05	327.2
1889	0.04	294.0	1931	0.03	307.7	1973	0.18	328.8
1890	−0.22	294.2	1932	0.04	308.2	1974	−0.06	329.7
1891	−0.55	294.3	1933	−0.11	308.6	1975	−0.02	330.7
1892	−0.40	294.5	1934	0.05	309.0	1976	−0.21	331.8
1893	−0.39	294.6	1935	−0.08	309.4	1977	0.16	333.3
1894	−0.32	294.7	1936	0.01	309.8	1978	0.07	334.6
1895	−0.32	294.8	1937	0.12	310.0	1979	0.13	336.9
1896	−0.27	294.9	1938	0.15	310.2	1980	0.27	338.7
1897	−0.15	295.0	1939	−0.02	310.3	1981	0.40	339.9
1898	−0.21	295.2	1940	0.14	310.4	1982	0.10	341.1
1899	−0.25	295.5	1941	0.11	310.4	1983	0.34	342.8
1900	−0.05	295.8	1942	0.10	310.3	1984	0.16	344.4
1901	−0.05	296.1	1943	0.06	310.2	1985	0.13	345.9
1902	−0.30	296.5	1944	0.10	310.1	1986	0.19	347.2
1903	−0.35	296.8	1945	−0.01	310.1	1987	0.35	348.9
1904	−0.42	297.2	1946	0.01	310.1	1988	0.42	351.5
1905	−0.25	297.6	1947	0.12	310.2	1989	0.28	352.9
1906	−0.15	298.1	1948	−0.03	310.3	1990	0.49	354.2
1907	−0.41	298.5	1949	−0.09	310.5	1991	0.44	355.6
1908	−0.30	298.9	1950	−0.17	310.7	1992	0.16	356.4
1909	−0.31	299.3	1951	−0.02	311.1	1993	0.18	357.0
1910	−0.21	299.7	1952	0.03	311.5	1994	0.31	358.9
1911	−0.25	300.1	1953	0.12	311.9	1995	0.47	360.9
1912	−0.33	300.4	1954	−0.09	312.4	1996	0.36	362.6
1913	−0.28	300.8	1955	−0.09	313.0	1997	0.40	363.8
1914	−0.02	301.1	1956	−0.18	313.6	1998	0.71	366.6
1915	0.06	301.4	1957	0.08	314.2	1999	0.43	368.3
1916	−0.20	301.7	1958	0.10	314.9	2000	0.41	369.5
1917	−0.46	302.1	1959	0.05	315.8	2001	0.56	371.0
1918	−0.33	302.4	1960	−0.02	316.6	2002	0.70	373.1
1919	−0.09	302.7	1961	0.10	317.3	2003	0.66	375.6
1920	−0.15	303.0	1962	0.05	318.1	2004	0.60	377.4
1921	−0.04	303.4	1963	0.03	318.7			

Exercises for Section 6.6

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

6.6.1 SS VS An article in the *Journal of Pharmaceuticals Sciences* ["Statistical Analysis of the Extended Hansen Method Using the Bootstrap Technique" (1991, Vol. 80, pp. 971–977)] presented data on the observed mole fraction solubility of a solute at a constant temperature and the dispersion, dipolar, and hydrogen-bonding Hansen partial solubility parameters. The data are as shown in the table below, where y is the negative logarithm of the mole fraction solubility, x_1 is the dispersion partial solubility, x_2 is the dipolar partial solubility, and x_3 is the hydrogen bonding partial solubility.

a. Construct a matrix of scatter plots for these variables.

b. Comment on the apparent relationships among y and the other three variables.

6.6.2 Consider the petroleum imports in Exercise 6.5.2.

a. Construct a matrix of scatter plots for the three variables and compute the correlation coefficients.

b. Comment on the apparent relationships among the variables.

Observation Number	y	x_1	x_2	x_3
1	0.22200	7.3	0.0	0.0
2	0.39500	8.7	0.0	0.3
3	0.42200	8.8	0.7	1.0
4	0.43700	8.1	4.0	0.2
5	0.42800	9.0	0.5	1.0
6	0.46700	8.7	1.5	2.8
7	0.44400	9.3	2.1	1.0
8	0.37800	7.6	5.1	3.4
9	0.49400	10.0	0.0	0.3
10	0.45600	8.4	3.7	4.1
11	0.45200	9.3	3.6	2.0
12	0.11200	7.7	2.8	7.1
13	0.43200	9.8	4.2	2.0
14	0.10100	7.3	2.5	6.8
15	0.23200	8.5	2.0	6.6
16	0.30600	9.5	2.5	5.0
17	0.09230	7.4	2.8	7.8
18	0.11600	7.8	2.8	7.7
19	0.07640	7.7	3.0	8.0
20	0.43900	10.3	1.7	4.2
21	0.09440	7.8	3.3	8.5
22	0.11700	7.1	3.9	6.6
23	0.07260	7.7	4.3	9.5
24	0.04120	7.4	6.0	10.9
25	0.25100	7.3	2.0	5.2
26	0.00002	7.6	7.8	20.7

Exercises for Section 6.7

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

6.7.1 WP Construct a normal probability plot of the visual accommodation data in Exercise 6.1.7. Does it seem reasonable to assume that visual accommodation is normally distributed?

6.7.2 WP Construct a normal probability plot of the octane rating data in Exercise 6.2.4. Does it seem reasonable to assume that octane rating is normally distributed?

6.7.3 WP SS VS Construct a normal probability plot of the cycles to failure data in Exercise 6.2.5. Does it seem reasonable to assume that cycles to failure are normally distributed?

6.7.4 Construct a normal probability plot of the suspended solids concentration data in Exercise 6.2.10. Does it seem reasonable to assume that the concentration of suspended solids in water from this particular lake is normally distributed?

6.7.5 Construct two normal probability plots for the height data in Exercises 6.2.8 and 6.2.13. Plot the data for female and male students on the same axes. Does height seem to be normally distributed for either group of students? If both populations have the same variance, the two normal probability plots should have identical slopes. What conclusions would you draw about the heights of the two groups of students from visual examination of the normal probability plots?

6.7.6 It is possible to obtain a "quick-and-dirty" estimate of the mean of a normal distribution from the 50th percentile value on a normal probability plot. Provide an argument why this is so. It is also possible to obtain an estimate of the standard deviation of a normal distribution by subtracting the 84th percentile value from the 50th percentile value. Provide an argument explaining why this is so.

Supplemental Exercises for Chapter 6

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

6.S7 **WP** **SS** The table below shows unemployment data for the United States that are seasonally adjusted. Construct a time series plot of these data and comment on any features (*Source:* U.S. Bureau of Labor Web site, http://data.bls.gov).

6.S8 A sample of six resistors yielded the following resistances (ohms): $x_1 = 45$, $x_2 = 38$, $x_3 = 47$, $x_4 = 41$, $x_5 = 35$, and $x_6 = 43$.

a. Compute the sample variance and sample standard deviation.

b. Subtract 35 from each of the original resistance measurements and compute s^2 and s. Compare your results with those obtained in part (a) and explain your findings.

c. If the resistances were 450, 380, 470, 410, 350, and 430 ohms, could you use the results of previous parts of this problem to find s^2 and s?

6.S9 **WP** Consider the following two samples:

Sample 1: 10, 9, 8, 7, 8, 6, 10, 6
Sample 2: 10, 6, 10, 6, 8, 10, 8, 6

a. Calculate the sample range for both samples. Would you conclude that both samples exhibit the same variability? Explain.

b. Calculate the sample standard deviations for both samples. Do these quantities indicate that both samples have the same variability? Explain.

c. Write a short statement contrasting the sample range versus the sample standard deviation as a measure of variability.

6.S10 **WP** The following data are the temperatures of effluent at discharge from a sewage treatment facility on consecutive days:

43	47	51	48	52	50	46	49
45	52	46	51	44	49	46	51
49	45	44	50	48	50	49	50

a. Calculate the sample mean, sample median, sample variance, and sample standard deviation.

b. Construct a box plot of the data and comment on the information in this display.

6.S11 **WP** **SS** **VS** The total net electricity consumption of the United States by year from 1980 to 2007 (in billion kilowatt-hours) is in the table below (*Source:* U.S. Department of Energy Web site, www.eia.doe.gov/emeu/international/contents.html#InternationalElectricity). Net consumption excludes the energy consumed by the generating units.

1980	2094.4	1981	2147.1	1982	2086.4	1983	2151.0
1984	2285.8	1985	2324.0	1986	2368.8	1987	2457.3
1988	2578.1	1989	2755.6	1990	2837.1	1991	2886.1
1992	2897.2	1993	3000.7	1994	3080.9	1995	3164.0
1996	3253.8	1997	3301.8	1998	3425.1	1999	3483.7
2000	3592.4	2001	3557.1	2002	3631.7	2003	3662.0
2004	3715.9	2005	3811.0	2006	3816.8	2007	3891.7

Construct a time series plot of these data. Construct and interpret a stem-and-leaf display of these data.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1999	4.3	4.4	4.2	4.3	4.2	4.3	4.3	4.2	4.2	4.1	4.1	4.0
2000	4.0	4.1	4.0	3.8	4.0	4.0	4.0	4.1	3.9	3.9	3.9	3.9
2001	4.2	4.2	4.3	4.4	4.3	4.5	4.6	4.9	5.0	5.3	5.5	5.7
2002	5.7	5.7	5.7	5.9	5.8	5.8	5.8	5.7	5.7	5.7	5.9	6.0
2003	5.8	5.9	5.9	6.0	6.1	6.3	6.2	6.1	6.1	6.0	5.8	5.7
2004	5.7	5.6	5.8	5.6	5.6	5.6	5.5	5.4	5.4	5.5	5.4	5.4
2005	5.2	5.4	5.2	5.2	5.1	5.1	5.0	4.9	5.0	5.0	5.0	4.8
2006	4.7	4.8	4.7	4.7	4.7	4.6	4.7	4.7	4.5	4.4	4.5	4.4
2007	4.6	4.5	4.4	4.5	4.5	4.6	4.7	4.7	4.7	4.8	4.7	4.9
2008	4.9	4.8	5.1	5.0	5.5	5.6	5.8	6.2	6.2	6.6	6.8	7.2
2009	7.6	8.1	8.5	8.9	9.4	9.5	9.4	9.7	9.8			

6.S12 Transformations. In some data sets, a transformation by some mathematical function applied to the original data, such as $\sqrt{y}$ or $\log y$, can result in data that are simpler to work with statistically than the original data. To illustrate the effect of a transformation, consider the following data, which represent cycles to failure for a yarn product: 675, 3650, 175, 1150, 290, 2000, 100, 375.

a. Construct a normal probability plot and comment on the shape of the data distribution.

b. Transform the data using logarithms; that is, let y^*(new value) = log y(old value). Construct a normal probability plot of the transformed data and comment on the effect of the transformation.

6.S13 In 1879, A. A. Michelson made 100 determinations of the velocity of light in air using a modification of a method proposed by French physicist Léon Foucault. Michelson made the measurements in five trials of 20 measurements each. The observations (in kilometers per second) are in the table below. Each value has 299,000 subtracted from it.

The currently accepted true velocity of light in a vacuum is 299,792.5 kilometers per second. Stigler ["Do Robust Estimators Work with Real Data?" *Annals of Statistics* (1977, Vol. 5(6), pp. 1055–1098)] reported that the "true" value for comparison to these measurements is 734.5 because of air effects. Construct comparative box plots of these measurements. Does it seem that all five trials are consistent with respect to the variability of the measurements? Are all five trials centered on the same value? How does each group of trials compare to the true value? Could there have been "startup" effects in the experiment that Michelson performed? Could there have been bias in the measuring instrument?

Trial 1

850	900	930	950	980
1000	930	760	1000	960
740	1070	850	980	880
980	650	810	1000	960

Trial 2

960	960	880	850	900
830	810	880	800	760
940	940	800	880	840
790	880	830	790	800

Trial 3

880	880	720	620	970
880	850	840	850	840
880	860	720	860	950
910	870	840	840	840

Trial 4

890	810	800	760	750
910	890	880	840	850
810	820	770	740	760
920	860	720	850	780

Trial 5

890	780	760	790	820
870	810	810	950	810
840	810	810	810	850
870	740	940	800	870

6.S14 WP In 1789, Henry Cavendish estimated the density of the Earth by using a torsion balance. His 29 measurements follow, expressed as a multiple of the density of water.

5.50	5.30	5.47	5.10	5.29	5.65
5.55	5.61	5.75	5.63	5.27	5.44
5.57	5.36	4.88	5.86	5.34	5.39
5.34	5.53	5.29	4.07	5.85	5.46
5.42	5.79	5.62	5.58	5.26	

a. Calculate the sample mean, sample standard deviation, and median of the Cavendish density data.

b. Construct a normal probability plot of the data. Comment on the plot. Does there seem to be a "low" outlier in the data?

c. Would the sample median be a better estimate of the density of the earth than the sample mean? Why?

6.S15 In their book *Introduction to Time Series Analysis and Forecasting* (Wiley, 2008), Montgomery, Jennings, and Kulahci presented the data on the drowning rate for children between one and four years old per 100,000 of population in Arizona from 1970 to 2004. The data are: 19.9, 16.1, 19.5, 19.8, 21.3, 15.0, 15.5, 16.4, 18.2, 15.3, 15.6, 19.5, 14.0, 13.1, 10.5, 11.5, 12.9, 8.4, 9.2, 11.9, 5.8, 8.5, 7.1, 7.9, 8.0, 9.9, 8.5, 9.1, 9.7, 6.2, 7.2, 8.7, 5.8, 5.7, and 5.2.

a. Perform an appropriate graphical analysis of the data.

b. Calculate and interpret the appropriate numerical summaries.

c. Notice that the rate appears to decrease dramatically starting about 1990. Discuss some potential reasons explaining why this could have happened.

d. If there has been a real change in the drowning rate beginning about 1990, what impact does this have on the summary statistics that you calculated in part (b)?

6.S16 The energy consumption for 90 gas-heated homes during a winter heating season is given in the table below. The variable reported is BTU/number of heating degree days.

a. Calculate the sample mean and standard deviation of energy usage.

b. Construct a histogram of the energy usage data and comment on the shape of the data distribution.

c. Construct a stem-and-leaf diagram of energy usage.

d. What proportion of the energy usage data is above the average usage plus 2 standard deviations?

7.87	9.43	7.16	8.67	12.31	9.84	16.90	10.04	12.62	7.62
11.12	13.43	9.07	6.94	10.28	9.37	7.93	13.96	6.80	4.00
8.58	8.00	5.98	15.24	8.54	11.09	11.70	12.71	6.78	9.82
12.91	10.35	9.60	9.58	9.83	9.52	18.26	10.64	6.62	5.20
12.28	7.23	2.97	8.81	9.27	11.29	8.29	9.96	10.30	16.06
14.24	11.43	10.28	13.60	5.94	10.36	6.85	6.72	10.21	8.61
11.62	11.21	10.95	7.62	10.40	12.92	15.12	13.47	8.47	11.70
7.73	8.37	7.29	10.49	8.69	8.26	7.69	12.19	5.56	9.76
7.15	12.69	13.38	13.11	10.50	14.35	13.42	6.35	9.83	12.16

6.S17 One of the authors (DCM) has a Mercedes-Benz 500 SL Roadster. It is a 2003 model and has fairly low mileage (currently 60,112 miles on the odometer). He is interested in learning how his car's mileage compares with the mileage on similar SLs. The table below contains the mileage on 100 Mercedes-Benz SLs from the model years 2003–2009 taken from the Cars.com website.

a. Calculate the sample mean and standard deviation of the odometer readings.

b. Construct a histogram of the odometer readings and comment on the shape of the data distribution.

c. Construct a stem-and-leaf diagram of the odometer readings.

d. What is the percentile of DCM's mileage?

6.S18 Consider the global mean surface air temperature anomaly and the global CO_2 concentration data originally shown in Exercise 6.5.4.

a. Construct a scatter plot of the global mean surface air temperature anomaly versus the global CO_2 concentration. Comment on the plot.

b. What is the simple correlation coefficient between these two variables?

2020	8905	1698	17,971	6207	22,643	4977	17,656	8940	11,508
7893	10,327	37,687	15,000	4166	9056	19,842	15,598	33,745	22,168
19,000	31,668	33,512	28,522	5824	18,327	31,845	30,015	2171	36,161
15,984	16,903	37,789	28,958	40,944	18,498	40,057	15,272	28,968	30,487
32,271	36,889	21,564	31,000	42,915	19,377	19,634	26,313	43,049	30,396
38,277	72,272	3800	21,218	29,250	48,648	29,216	44,944	49,125	33,065
32,524	38,139	62,940	51,326	54,126	4100	45,540	26,235	46,505	34,420
15,972	41,218	43,382	15,879	13,500	77,809	25,708	29,000	58,006	51,071
63,249	58,526	66,325	49,489	32,800	67,000	60,499	63,260	60,449	27,422
60,583	83,500	56,314	67,072	62,500	47,603	51,936	65,195	64,473	85,475

Chapter 7 Exercises

Exercises for Section 7.2

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

7.2.1 **WP** Consider the compressive strength data in Table 6.2. What proportion of the specimens exhibit compressive strength of at least 200 psi? What is the population parameter estimated here?

7.2.2 **WP** Consider the synthetic fiber in the previous exercise. How is the standard deviation of the sample mean changed when the sample size is increased from $n = 6$ to $n = 49$?

7.2.3 **WP** **VS** A synthetic fiber used in manufacturing carpet has tensile strength that is normally distributed with mean 75.5 psi and standard deviation 3.5 psi. Find the probability that a random sample of $n = 6$ fiber specimens will have sample mean tensile strength that exceeds 75.75 psi.

7.2.4 WP Suppose that the random variable X has the continuous uniform distribution

$$f(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Suppose that a random sample of $n = 12$ observations is selected from this distribution. What is the approximate probability distribution of $\overline{X} - 6$? Find the mean and variance of this quantity.

7.2.5 WP A normal population has mean 100 and variance 25. How large must the random sample be if you want the standard error of the sample average to be 1.5?

7.2.6 WP VS The amount of time that a customer spends waiting at an airport check-in counter is a random variable with mean 8.2 minutes and standard deviation 1.5 minutes. Suppose that a random sample of $n = 49$ customers is observed. Find the probability that the average time waiting in line for these customers is

a. Less than 10 minutes **b.** Between 5 and 10 minutes

c. Less than 6 minutes

7.2.7 WP SS VS A random sample of size $n_1 = 16$ is selected from a normal population with a mean of 75 and a standard deviation of 8. A second random sample of size $n_2 = 9$ is taken from another normal population with mean 70 and standard deviation 12. Let $\overline{X}_1$ and $\overline{X}_2$ be the two sample means. Find:

a. The probability that $\overline{X}_1 - \overline{X}_2$ exceeds 4

b. The probability that $3.5 \le \overline{X}_1 - \overline{X}_2 \le 5.5$

7.2.8 Scientists at the Hopkins Memorial Forest in western Massachusetts have been collecting meteorological and environmental data in the forest data for more than 100 years. In the past few years, sulfate content in water samples from Birch Brook has averaged 7.48 mg/L with a standard deviation of 1.60 mg/L.

a. What is the standard error of the sulfate in a collection of 10 water samples?

b. If 10 students measure the sulfate in their samples, what is the probability that their average sulfate will be between 6.49 and 8.47 mg/L?

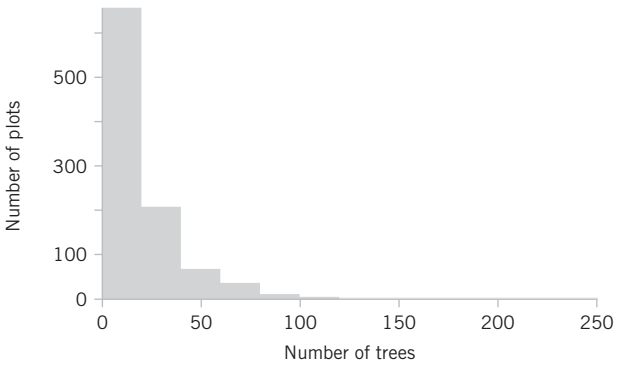
c. What do you need to assume for the probability calculated in (b) to be accurate?

7.2.9 Data on the pH of rain in Ingham County, Michigan, are as follows:

5.47 5.37 5.38 4.63 5.37 3.74 3.71 4.96 4.64 5.11 5.65 5.39 4.16 5.62 4.57 4.64 5.48 4.57 4.57 4.51 4.86 4.56 4.61 4.32 3.98 5.70 4.15 3.98 5.65 3.10 5.04 4.62 4.51 4.34 4.16 4.64 5.12 3.71 4.64

What proportion of the samples has pH below 5.0?

7.2.10 Researchers in the Hopkins Forest (see Exercise 7.2.8) also count the number of maple trees (genus *acer*) in plots throughout the forest. The following is a histogram of the number of live maples in 1002 plots sampled over the past 20 years. The average number of maples per plot was 19.86 trees with a standard deviation of 23.65 trees.



a. If we took the mean of a sample of eight plots, what would be the standard error of the mean?

b. Using the central limit theorem, what is the probability that the mean of the eight would be within 1 standard error of the mean?

c. Why might you think that the probability that you calculated in (b) might not be very accurate?

7.2.11 Like hurricanes and earthquakes, geomagnetic storms are natural hazards with possible severe impact on the earth. Severe storms can cause communication and utility breakdowns, leading to possible blackouts. The National Oceanic and Atmospheric Administration beams electron and proton flux data in various energy ranges to various stations on the earth to help forecast possible disturbances. The following are 25 readings of proton flux in the 47-68 kEV range (units are in p/(cm2-sec-ster-MeV)) on the evening of December 28, 2011:

2310 2320 2010 10800 2190 3360 5640 2540 3360 11800 2010 3430 10600 7370 2160 3200 2020 2850 3500 10200 8550 9500 2260 7730 2250

a. Find a point estimate of the mean proton flux in this time period.

b. Find a point estimate of the standard deviation of the proton flux in this time period.

c. Find an estimate of the standard error of the estimate in part (a).

d. Find a point estimate for the median proton flux in this time period.

e. Find a point estimate for the proportion of readings that are less than 5000 p/(cm2-sec-ster-MeV).

7.2.12 Wayne Collier designed an experiment to measure the fuel efficiency of his family car under different tire pressures. For each run, he set the tire pressure and then measured the miles he drove on a highway (I-95 between Mills River and Pisgah Forest, NC) until he ran out of fuel using 2 liters of fuel each time. To do this, he made some alterations to the normal flow of gasoline to the engine. In Wayne's words, "I inserted a T-junction into the fuel line just before the fuel filter, and a line into the passenger compartment of my car, where it joined with a graduated 2-liter Rubbermaid© bottle that I mounted in a box where the passenger seat is normally fastened. Then I sealed off the fuel-return line, which under normal operation sends excess fuel from the fuel pump back to the fuel tank."

Suppose that you call the mean miles that he can drive with normal pressure in the tires μ. An unbiased estimate for μ is the mean of the sample runs, $\bar{x}$. But Wayne has a different idea. He decides to use the following estimator: He flips a fair coin. If the coin comes up heads, he will add 5 miles to each observation. If tails come up, he will subtract 5 miles from each observation.

a. Show that Wayne's estimate is, in fact, unbiased.
b. Compare the standard deviation of Wayne's estimate with the standard deviation of the sample mean.
c. Given your answer to (b), why does Wayne's estimate not make good sense scientifically?

Exercises for Section 7.3

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

7.3.1 A computer software package calculated some numerical summaries of a sample of data. The results are displayed here:

Variable	N	Mean	SE Mean	StDev	Variance	Sum	Sum of Squares
x	?	?	2.05	10.25	?	3761.70	?

a. Fill in the missing quantities.
b. What is the estimate of the mean of the population from which this sample was drawn?

7.3.2 Suppose we have a random sample of size $2n$ from a population denoted by X, and $E(X) = \mu$ and $V(X) = \sigma^2$. Let

$$\bar{X}_1 = \frac{1}{2n}\sum_{i=1}^{2n} X_i \quad \text{and} \quad \bar{X}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

be two estimators of μ. Which is the better estimator of μ? Explain your choice.

7.3.3 **SS** Suppose that we have a random sample $X_1, X_2, \ldots, X_n$ from a population that is $N(\mu, \sigma^2)$. We plan to use $\hat{\Theta} = \sum_{i=1}^{n}(X_i - \bar{X})^2/c$ to estimate σ^2. Compute the bias in $\hat{\Theta}$ as an estimator of σ^2 as a function of the constant c.

7.3.4 **WP** **VS** Suppose that $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are unbiased estimators of the parameter θ. We know that $V(\hat{\Theta}_1) = 10$ and $V(\hat{\Theta}_2) = 4$. Which estimator is better and in what sense is it better? Calculate the relative efficiency of the two estimators.

7.3.5 Suppose that $\hat{\Theta}_1, \hat{\Theta}_2$, and $\hat{\Theta}_3$ are estimators of θ. We know that $E(\hat{\Theta}_1) = E(\hat{\Theta}_2) = \theta$, $E(\hat{\Theta}_3) \neq \theta$, $V(\hat{\Theta}_1) = 12$, $V(\hat{\Theta}_2) = 10$, and $E(\hat{\Theta}_3 - \theta)^2 = 6$. Compare these three estimators. Which do you prefer? Why?

7.3.6 **WP** Data on pull-off force (pounds) for connectors used in an automobile engine application are as follows: 79.3, 75.1, 78.2, 74.1, 73.9, 75.0, 77.6, 77.3, 73.8, 74.6, 75.5, 74.0, 74.7, 75.9, 72.9, 73.8, 74.2, 78.1, 75.4, 76.3, 75.3, 76.2, 74.9, 78.0, 75.1, 76.8.

a. Calculate a point estimate of the mean pull-off force of all connectors in the population. State which estimator you used and why.
b. Calculate a point estimate of the pull-off force value that separates the weakest 50% of the connectors in the population from the strongest 50%.
c. Calculate point estimates of the population variance and the population standard deviation.
d. Calculate the standard error of the point estimate found in part (a). Interpret the standard error.
e. Calculate a point estimate of the proportion of all connectors in the population whose pull-off force is less than 73 pounds.

7.3.7 **WP** **SS** **VS** Data on the oxide thickness of semiconductor wafers are as follows: 425, 431, 416, 419, 421, 436, 418, 410, 431, 433, 423, 426, 410, 435, 436, 428, 411, 426, 409, 437, 422, 428, 413, 416.

a. Calculate a point estimate of the mean oxide thickness for all wafers in the population.
b. Calculate a point estimate of the standard deviation of oxide thickness for all wafers in the population.
c. Calculate the standard error of the point estimate from part (a).
d. Calculate a point estimate of the median oxide thickness for all wafers in the population.
e. Calculate a point estimate of the proportion of wafers in the population that have oxide thickness of more than 430 angstroms.

7.3.8 Suppose that X is the number of observed "successes" in a sample of n observations where p is the probability of success on each observation.

a. Show that $\hat{P} = X/n$ is an unbiased estimator of p.
b. Show that the standard error of $\hat{P}$ is $\sqrt{p(1-p)/n}$. How would you estimate the standard error?

7.3.9 **VS** $\bar{X}_1$ and S_1^2 are the sample mean and sample variance from a population with mean μ_1 and variance σ_1^2. Similarly, $\bar{X}_2$ and S_2^2 are the sample mean and sample variance from a second independent population with mean μ_2 and variance σ_2^2. The sample sizes are n_1 and n_2, respectively.

a. Show that $\bar{X}_1 - \bar{X}_2$ is an unbiased estimator of $\mu_1 - \mu_2$.
b. Find the standard error of $\bar{X}_1 - \bar{X}_2$. How could you estimate the standard error?
c. Suppose that both populations have the same variance; that is, $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Show that

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

is an unbiased estimator of σ^2.

7.3.10 Consider a normal random variable with mean 10 and standard deviation 4. Suppose that a random sample of size 16 is drawn from this distribution and the sample mean is computed.

We know that the standard error of the sample mean in this case is $\sigma/\sqrt{n} = \sigma/\sqrt{16} = 1$. Use the bootstrap method with $n_B = 200$ bootstrap samples to find the standard error of the sample mean. Compare the bootstrap standard error to the actual standard error.

7.3.11 Of n_1 randomly selected engineering students at Arizona State University, X_1 owned an Apple computer, and of n_2 randomly selected engineering students at Virginia Tech, X_2 owned an Apple computer. Let p_1 and p_2 be the probability that randomly selected ASU and Virginia Tech engineering students, respectively, own Apple computers.

a. Show that an unbiased estimate for $p_1 - p_2$ is $(X_1/n_1) - (X_2/n_2)$.

b. What is the standard error of the point estimate in part (a)?

c. How would you compute an estimate of the standard error found in part (b)?

d. Suppose that $n_1 = 200$, $X_1 = 150$, $n_2 = 250$, and $X_2 = 185$. Use the results of part (a) to compute an estimate of $p_1 - p_2$.

e. Use the results in parts (b) through (d) to compute an estimate of the standard error of the estimate.

7.3.12 Suppose that two independent random samples (of size n_1 and n_2) from two normal distributions are available. Explain how you would estimate the standard error of the difference in sample means $\overline{X}_1 - \overline{X}_2$ with the bootstrap method.

Exercises for Section 7.4

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

7.4.1 **WP** **SS** **VS** Let X be a geometric random variable with parameter p. Find the maximum likelihood estimator of p based on a random sample of size n.

7.4.2 **WP** **VS** Consider the probability density function

$$f(x) = \frac{1}{\theta^2} x e^{-x/\theta}, \qquad 0 \le x < \infty, \quad 0 < \theta < \infty$$

Find the maximum likelihood estimator for θ.

7.4.3 Let $X_1, X_2, \ldots, X_n$ be uniformly distributed on the interval 0 to a. Show that the moment estimator of a is $\hat{a} = 2\overline{X}$. Is this an unbiased estimator? Discuss the reasonableness of this estimator.

7.4.4 Consider the probability density function

$$f(x) = c(1 + \theta x), \quad -1 \le x \le 1$$

a. Find the value of the constant c.

b. What is the moment estimator for θ?

c. Show that $\hat{\theta} = 3\overline{X}$ is an unbiased estimator for θ.

d. Find the maximum likelihood estimator for θ.

7.4.5 Consider the Weibull distribution

$$f(x) = \begin{cases} \dfrac{\beta}{\delta}\left(\dfrac{x}{\delta}\right)^{\beta-1} e^{-\left(\frac{x}{\delta}\right)^{\beta}}, & 0 < x \\ 0, & \text{otherwise} \end{cases}$$

a. Find the likelihood function based on a random sample of size n. Find the log likelihood.

b. Show that the log likelihood is maximized by solving the following equations

$$\beta = \left[\frac{\sum_{i=1}^{n} x_i^{\beta} \ln(x_i)}{\sum_{i=1}^{n} x_i^{\beta}} - \frac{\sum_{i=1}^{n} \ln(x_i)}{n} \right]^{-1} \qquad \delta = \left[\frac{\sum_{i=1}^{n} x_i^{\beta}}{n} \right]^{1/\beta}$$

c. What complications are involved in solving the two equations in part (b)?

7.4.6 Let $X_1, X_2, \ldots, X_n$ be uniformly distributed on the interval 0 to a. Recall that the maximum likelihood estimator of a is $\hat{a} = \max(X_i)$.

a. Argue intuitively why $\hat{a}$ cannot be an unbiased estimator for a.

b. Suppose that $E(\hat{a}) = na/(n+1)$. Is it reasonable that $\hat{a}$ consistently underestimates a? Show that the bias in the estimator approaches zero as n gets large.

c. Propose an unbiased estimator for a.

d. Let $Y = \max(X_i)$. Use the fact that $Y \le y$ if and only if each $X_i \le y$ to derive the cumulative distribution function of Y. Then show that the probability density function of Y is

$$f(y) = \begin{cases} \dfrac{n y^{n-1}}{a^n}, & 0 \le y \le a \\ 0, & \text{otherwise} \end{cases}$$

Use this result to show that the maximum likelihood estimator for a is biased.

e. We have two unbiased estimators for a: the moment estimator $\hat{a}_1 = 2\overline{X}$ and $\hat{a}_2 = [(n+1)/n]\max(X_i)$, where $\max(X_i)$ is the largest observation in a random sample of size n. It can be shown that $V(\hat{a}_1) = a^2/(3n)$ and that $V(\hat{a}_2) = a^2/[n(n+2)]$. Show that if $n > 1$, $\hat{a}_2$ is a better estimator than $\hat{a}$. In what sense is it a better estimator of a?

7.4.7 SS Reconsider the oxide thickness data in Exercise 7.3.7 and suppose that it is reasonable to assume that oxide thickness is normally distributed.

a. Compute the maximum likelihood estimates of μ and σ^2.

b. Graph the likelihood function in the vicinity of $\hat{\mu}$ and $\hat{\sigma}^2$, the maximum likelihood estimates, and comment on its shape.

c. Suppose that the sample size was larger ($n = 40$) but the maximum likelihood estimates were numerically equal to the values obtained in part (a). Graph the likelihood function for $n = 40$, compare it to the one from part (b), and comment on the effect of the larger sample size.

Supplemental Exercises for Chapter 7

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutoring Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

7.S8 Suppose that a random variable is normally distributed with mean μ and variance σ^2, and we draw a random sample of five observations from this distribution. What is the joint probability density function of the sample?

7.S9 WP Transistors have a life that is exponentially distributed with parameter λ. A random sample of n transistors is taken. What is the joint probability density function of the sample?

7.S10 WP VS A procurement specialist has purchased 25 resistors from vendor 1 and 30 resistors from vendor 2. Let $X_{1,1}$, $X_{1,2}, \ldots, X_{1,25}$ represent the vendor 1 observed resistances, which are assumed to be normally and independently distributed with mean 100 ohms and standard deviation 1.5 ohms. Similarly, let $X_{2,1}, X_{2,2}, \ldots, X_{2,30}$ represent the vendor 2 observed resistances, which are assumed to be normally and independently distributed with mean 105 ohms and standard deviation of 2.0 ohms. What is the sampling distribution of $\overline{X}_1 - \overline{X}_2$? What is the standard error of $\overline{X}_1 - \overline{X}_2$?

7.S11 WP SS VS A random sample of 36 observations has been drawn from a normal distribution with mean 50 and standard deviation 12. Find the probability that the sample mean is in the interval $47 \le X \le 53$. Is the assumption of normality important? Why?

7.S12 A manufacturer of semiconductor devices takes a random sample of 100 chips and tests them, classifying each chip as defective or nondefective. Let $X_i = 0$ if the chip is nondefective and $X_i = 1$ if the chip is defective. The sample fraction defective is

$$\hat{P} = \frac{X_1 + X_2 + \cdots + X_{100}}{100}$$

What is the sampling distribution of the random variable $\hat{P}$?

7.S13 WP A random variable x has probability density function

$$f(x) = \frac{1}{2\theta^3} x^2 e^{-x/\theta}, \qquad 0 < x < \infty, \quad 0 < \theta < \infty$$

Find the maximum likelihood estimator for θ.

7.S14 Let $f(x) = (1/\theta)x^{(1-\theta)/\theta}$, $0 < x < 1$, and $0 < \theta < \infty$. Show that $\hat{\Theta} = -(1/n)\sum_{i=1}^{n} \ln(X_i)$ is the maximum likelihood estimator for θ and that $\hat{\Theta}$ is an unbiased estimator for q.

7.S15 An electric utility has placed special meters on 10 houses in a subdivision that measures the energy consumed (demand) at each hour of the day. The company is interested in the energy demand at one specific hour—the hour at which the system experiences peak consumption. The data from these 10 meters are as follows (in KW): 23.1, 15.6, 17.4, 20.1, 19.8, 26.4, 25.1, 20.5, 21.9, and 28.7. If μ is the true mean peak demand for the 10 houses in this group of houses having the special meters, estimate μ. Now suppose that the utility wants to estimate the demand at the peak hour for all 5000 houses in this subdivision. Let θ be this quantity. Estimate θ using the data given. Estimate the proportion of houses in the subdivision that demand at least 20 KW at the hour of system peak.

Chapter 8 Exercises

Exercises for Section 8.1

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutoring Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

8.1.1 WP For a normal population with known variance σ^2, answer the following questions:

a. What is the confidence level for the interval $\overline{x} - 2.14\sigma/\sqrt{n} \le \mu \le \overline{x} + 2.14\sigma/\sqrt{n}$?

b. What is the confidence level for the interval $\bar{x} - 2.49\sigma/\sqrt{n} \bar{x} - 2.49\sigma/\sqrt{n} \le \mu \le \bar{x} + 2.49\sigma/\sqrt{n}$?

c. What is the confidence level for the interval $\bar{x} - 1.85\sigma/\sqrt{n} \le \mu \le \bar{x} + 1.85\sigma/\sqrt{n}$?

d. What is the confidence level for the interval $\mu \le \bar{x} + 2.00\sigma/\sqrt{n}$?

e. What is the confidence level for the interval $\bar{x} - 1.96\sigma/\sqrt{n} \le \mu$?

8.1.2 WP Consider the one-sided confidence interval expressions for a mean of a normal population.

a. What value of z_α would result in a 90% CI?

b. What value of z_α would result in a 95% CI?

c. What value of z_α would result in a 99% CI?

8.1.3 A random sample has been taken from a normal distribution and the following confidence intervals constructed using the same data: (38.02, 61.98) and (39.95, 60.05)

a. What is the value of the sample mean?

b. One of these intervals is a 95% CI and the other is a 90% CI. Which one is the 95% CI and why?

8.1.4 WP A confidence interval estimate is desired for the gain in a circuit on a semiconductor device. Assume that gain is normally distributed with standard deviation $\sigma = 20$.

a. How large must n be if the length of the 95% CI is to be 40?

b. How large must n be if the length of the 99% CI is to be 40?

8.1.5 Suppose that $n = 100$ random samples of water from a freshwater lake were taken and the calcium concentration (milligrams per liter) measured. A 95% CI on the mean calcium concentration is $0.49 \le \mu \le 0.82$.

a. Would a 99% CI calculated from the same sample data be longer or shorter?

b. Consider the following statement: There is a 95% chance that μ is between 0.49 and 0.82. Is this statement correct? Explain your answer.

c. Consider the following statement: If $n = 100$ random samples of water from the lake were taken and the 95% CI on μ computed, and this process were repeated 1000 times, 950 of the CIs would contain the true value of μ. Is this statement correct? Explain your answer.

8.1.6 WP SS The yield of a chemical process is being studied. From previous experience, yield is known to be normally distributed and $\sigma = 3$. The past 5 days of plant operation have resulted in the following percent yields: 91.6, 88.75, 90.8, 89.95, and 91.3. Find a 95% two-sided confidence interval on the true mean yield.

8.1.7 WP A manufacturer produces piston rings for an automobile engine. It is known that ring diameter is normally distributed with $\sigma = 0.001$ millimeters. A random sample of 15 rings has a mean diameter of $\bar{x} = 74.036$ millimeters.

a. Construct a 99% two-sided confidence interval on the mean piston ring diameter.

b. Construct a 99% lower-confidence bound on the mean piston ring diameter. Compare the lower bound of this confidence interval with the one in part (a).

8.1.8 WP VS A civil engineer is analyzing the compressive strength of concrete. Compressive strength is normally distributed with $\sigma^2 = 1000(\text{psi})^2$. A random sample of 12 specimens has a mean compressive strength of $\bar{x} = 3250$ psi.

a. Construct a 95% two-sided confidence interval on mean compressive strength.

b. Construct a 99% two-sided confidence interval on mean compressive strength. Compare the width of this confidence interval with the width of the one found in part (a).

8.1.9 WP Suppose that in Exercise 8.1.8 it is desired to estimate the compressive strength with an error that is less than 15 psi at 99% confidence. What sample size is required?

8.1.10 If the sample size n is doubled, by how much is the length of the CI on μ in Equation 8.5 reduced? What happens to the length of the interval if the sample size is increased by a factor of four?

8.1.11 WP SS By how much must the sample size n be increased if the length of the CI on μ in Equation 8.5 is to be halved?

8.1.12 Ishikawa et al. ["Evaluation of Adhesiveness of *Acinetobacter* sp. Tol 5 to Abiotic Surfaces," *Journal of Bioscience and Bioengineering* (Vol. 113(6), pp. 719–725)] studied the adhesion of various biofilms to solid surfaces for possible use in environmental technologies. Adhesion assay is conducted by measuring absorbance at A_{590}. Suppose that for the bacterial strain *Acinetobacter*, five measurements gave readings of 2.69, 5.76, 2.67, 1.62, and 4.12 dyne-cm^2. Assume that the standard deviation is known to be 0.66 dyne-cm^2.

a. Find a 95% confidence interval for the mean adhesion.

b. If the scientists want the confidence interval to be no wider than 0.55 dyne-cm^2, how many observations should they take?

8.1.13 WP GO Tutorial An article in the *Journal of Agricultural Science* ["The Use of Residual Maximum Likelihood to Model Grain Quality Characteristics of Wheat with Variety, Climatic and Nitrogen Fertilizer Effects" (1997, Vol. 128, pp. 135–142)] investigated means of wheat grain crude protein content (CP) and Hagberg falling number (HFN) surveyed in the United Kingdom. The analysis used a variety of nitrogen fertilizer applications (kg N/ha), temperature (°C), and total monthly rainfall (mm). The following data below describe temperatures for wheat grown at Harper Adams Agricultural College between 1982 and 1993. The temperatures measured in June were obtained as follows:

15.2	14.2	14.0	12.2	14.4	12.5
14.3	14.2	13.5	11.8	15.2	

Assume that the standard deviation is known to be $\sigma = 0.5$.

a. Construct a 99% two-sided confidence interval on the mean temperature.

b. Construct a 95% lower-confidence bound on the mean temperature.

c. Suppose that you wanted to be 95% confident that the error in estimating the mean temperature is less than 2 degrees Celsius. What sample size should be used?

d. Suppose that you wanted the total width of the two-sided confidence interval on mean temperature to be 1.5 degrees Celsius at 95% confidence. What sample size should be used?

Exercises for Section 8.2

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

8.2.1 A random sample has been taken from a normal distribution. Output from a software package follows:

Variable	N	Mean	SE Mean	StDev	Variance	Sum
x	10	?	0.507	1.605	?	251.848

a. Fill in the missing quantities.
b. Find a 95% CI on the population mean.

8.2.2 **WP** **SS** Find the values of the following percentiles: $t_{0.025,15}$, $t_{0.05,10}$, $t_{0.10,20}$, $t_{0.005,25}$, and $t_{0.001,30}$.

8.2.3 A research engineer for a tire manufacturer is investigating tire life for a new rubber compound and has built 16 tires and tested them to end-of-life in a road test. The sample mean and standard deviation are 60,139.7 and 3645.94 kilometers. Find a 95% confidence interval on mean tire life.

8.2.4 **WP** Determine the t-percentile that is required to construct each of the following one-sided confidence intervals:
a. Confidence level = 95%, degrees of freedom = 14
b. Confidence level = 99%, degrees of freedom = 19
c. Confidence level = 99.9%, degrees of freedom = 24

8.2.5 **WP** An article in *Obesity Research* ["Impaired Pressure Natriuresis in Obese Youths" (2003, Vol. 11, pp. 745–751)] described a study in which all meals were provided for 14 lean boys for three days followed by one stress test (with a video-game task). The average systolic blood pressure (SBP) during the test was 118.3 mm HG with a standard deviation of 9.9 mm HG. Construct a 99% one-sided upper confidence interval for mean SBP.

8.2.6 **WP** An article in *Medicine and Science in Sports and Exercise* ["Maximal Leg-Strength Training Improves Cycling Economy in Previously Untrained Men" (2005, Vol. 37, pp. 131–136)] studied cycling performance before and after 8 weeks of leg-strength training. Seven previously untrained males performed leg-strength training 3 days per week for 8 weeks (with four sets of five replications at 85% of one repetition maximum). Peak power during incremental cycling increased to a mean of 315 watts with a standard deviation of 16 watts. Construct a 95% confidence interval for the mean peak power after training.

8.2.7 The brightness of a television picture tube can be evaluated by measuring the amount of current required to achieve a particular brightness level. A sample of 10 tubes results in $\bar{x} = 317.2$ and $s = 15.7$. Find (in microamps) a 99% confidence interval on mean current required. State any necessary assumptions about the underlying distribution of the data.

8.2.8 An article in the *Journal of Composite Materials* (December 1989, Vol. 23(12), pp. 1200–1215) describes the effect of delamination on the natural frequency of beams made from composite laminates. Five such delaminated beams were subjected to loads, and the resulting frequencies (in hertz) were as follows:

230.66, 233.05, 232.58, 229.48, 232.58

Check the assumption of normality in the population. Calculate a 90% two-sided confidence interval on mean natural frequency.

8.2.9 The compressive strength of concrete is being tested by a civil engineer who tests 12 specimens and obtains the following data:

2216	2237	2249	2204
2225	2301	2281	2263
2318	2255	2275	2295

a. Check the assumption that compressive strength is normally distributed. Include a graphical display in your answer.
b. Construct a 95% two-sided confidence interval on the mean strength.
c. Construct a 95% lower confidence bound on the mean strength. Compare this bound with the lower bound of the two-sided confidence interval and discuss why they are different.

8.2.10 **WP** **SS** An article in *Computers & Electrical Engineering* ["Parallel Simulation of Cellular Neural Networks" (1996, Vol. 22, pp. 61–84)] considered the speedup of cellular neural networks (CNNs) for a parallel general-purpose computing architecture based on six transputers in different areas. The data follow:

3.775302	3.350679	4.217981	4.030324	4.639692
4.139665	4.395575	4.824257	4.268119	4.584193
4.930027	4.315973	4.600101		

a. Is there evidence to support the assumption that speedup of CNN is normally distributed? Include a graphical display in your answer.
b. Construct a 95% two-sided confidence interval on the mean speedup.
c. Construct a 95% lower confidence bound on the mean speedup.

8.2.11 An article in *Nuclear Engineering International* (February 1988, p. 33) describes several characteristics of fuel rods used in a reactor owned by an electric utility in Norway. Measurements on the percentage of enrichment of 12 rods were reported as follows:

2.94	3.00	2.90	2.75	3.00	2.95
2.90	2.75	2.95	2.82	2.81	3.05

a. Use a normal probability plot to check the normality assumption.
b. Find a 99% two-sided confidence interval on the mean percentage of enrichment. Are you comfortable with the statement that the mean percentage of enrichment is 2.95%? Why?

8.2.12 A healthcare provider monitors the number of CAT scans performed each month in each of its clinics. The most recent year of data for a particular clinic follows (the reported variable is the number of CAT scans each month expressed as the number of CAT scans per thousand members of the health plan):

2.31, 2.09, 2.36, 1.95, 1.98, 2.25, 2.16, 2.07, 1.88, 1.94, 1.97, 2.02

a. Find a 95% two-sided CI on the mean number of CAT scans performed each month at this clinic.
b. Historically, the mean number of scans performed by all clinics in the system has been 1.95. Is there any evidence that this particular clinic performs more CAT scans on average than the overall system average?

Exercises for Section 8.3

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

8.3.1 WP The percentage of titanium in an alloy used in aerospace castings is measured in 51 randomly selected parts. The sample standard deviation is $s = 0.37$. Construct a 95% two-sided confidence interval for σ.

8.3.2 WP VS Determine the values of the following percentiles:

$$\chi^2_{0.05,10}, \chi^2_{0.025,15}, \chi^2_{0.01,12}, \chi^2_{0.95,20}, \chi^2_{0.99,18}, \chi^2_{0.995,16}, \text{and } \chi^2_{0.005,25}$$

8.3.3 WP GO Tutorial VS SS An article in *Urban Ecosystems* ["Urbanization and Warming of Phoenix (Arizona, USA): Impacts, Feedbacks and Mitigation" (2002, Vol. 6, pp. 183–203)] mentions that Phoenix is ideal to study the effects of an urban heat island because it has grown from a population of 300,000 to approximately 3 million over the last 50 years, which is a period with a continuous, detailed climate record. The 50-year averages of the mean annual temperatures at eight sites in Phoenix follow. Check the assumption of normality in the population with a probability plot. Construct a 95% confidence interval for the standard deviation over the sites of the mean annual temperatures.

Site	Average Mean Temperature (°C)
Sky Harbor Airport	23.3
Phoenix Greenway	21.7
Phoenix Encanto	21.6
Waddell	21.7
Litchfield	21.3
Laveen	20.7
Maricopa	20.9
Harlquahala	20.1

8.3.4 WP An article in *Medicine and Science in Sports and Exercise* ["Electrostimulation Training Effects on the Physical Performance of Ice Hockey Players" (2005, Vol. 37, pp. 455–460)] considered the use of electromyostimulation (EMS) as a method to train healthy skeletal muscle. EMS sessions consisted of 30 contractions (4-second duration, 85 Hz) and were carried out three times per week for 3 weeks on 17 ice hockey players. The 10-meter skating performance test showed a standard deviation of 0.09 seconds. Construct a 95% confidence interval of the standard deviation of the skating performance test.

8.3.5 WP An article in *Technometrics* ["Two-Way Random Effects Analyses and Gauge R&R Studies" (1999, Vol. 41(3), pp. 202–211)] studied the capability of a gauge by measuring the weight of paper. The data for repeated measurements of one sheet of paper are in the following table. Construct a 95% one-sided upper confidence interval for the standard deviation of these measurements. Check the assumption of normality of the data and comment on the assumptions for the confidence interval.

Observations				
3.481	3.448	3.485	3.475	3.472
3.477	3.472	3.464	3.472	3.470
3.470	3.470	3.477	3.473	3.474

8.3.6 WP An article in *Cancer Research* ["Analyses of Litter-Matched Time-to-Response Data, with Modifications for Recovery of Interlitter Information" (1977, Vol. 37, pp. 3863–3868)] tested the tumorigenesis of a drug. Rats were randomly selected from litters and given the drug. The times of tumor appearance were recorded as follows:

101, 104, 104, 77, 89, 88, 104, 96, 82, 70, 89, 91, 39, 103, 93, 85, 104, 104, 81, 67, 104, 104, 104, 87, 104, 89, 78, 104, 86, 76, 103, 102, 80, 45, 94, 104, 104, 76, 80, 72, 73

Calculate a 95% confidence interval on the standard deviation of time until a tumor appearance. Check the assumption of normality of the population and comment on the assumptions for the confidence interval.

8.3.7 From the data on the pH of rain in Ingham County, Michigan:

5.47 5.37 5.38 4.63 5.37 3.74 3.71 4.96 4.64 5.11 5.65 5.39 4.16 5.62 4.57 4.64 5.48 4.57 4.57 4.51 4.86 4.56 4.61 4.32 3.98 5.70 4.15 3.98 5.65 3.10 5.04 4.62 4.51 4.34 4.16 4.64 5.12 3.71 4.64

Find a two-sided 95% confidence interval for the standard deviation of pH.

Exercises for Section 8.4

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

8.4.1 WP The 2004 presidential election exit polls from the critical state of Ohio provided the following results. The exit polls had 2020 respondents, 768 of whom were college graduates. Of the college graduates, 412 voted for George Bush.

a. Calculate a 95% confidence interval for the proportion of college graduates in Ohio who voted for George Bush.

b. Calculate a 95% lower confidence bound for the proportion of college graduates in Ohio who voted for George Bush.

8.4.2 WP VS An article in *Knee Surgery, Sports Traumatology, Arthroscopy* ["Arthroscopic Meniscal Repair with an Absorbable Screw: Results and Surgical Technique" (2005, Vol. 13, pp. 273–279)] showed that only 25 out of 37 tears (67.6%) located between 3 and 6 mm from the meniscus rim were healed.

a. Calculate a two-sided 95% confidence interval on the proportion of such tears that will heal.

b. Calculate a 95% lower confidence bound on the proportion of such tears that will heal.

8.4.3 WP An article in the *Journal of the American Statistical Association* ["Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling" (1990, Vol. 85, pp. 972–985)] measured the weight of 30 rats under experiment controls. Suppose that 12 were underweight rats.

a. Calculate a 95% two-sided confidence interval on the true proportion of rats that would show underweight from the experiment.

b. Using the point estimate of p obtained from the preliminary sample, what sample size is needed to be 95% confident that the error in estimating the true value of p is less than 0.02?

c. How large must the sample be if you wish to be at least 95% confident that the error in estimating p is less than 0.02, regardless of the true value of p?

8.4.4 WP SS The Arizona Department of Transportation wishes to survey state residents to determine what proportion of the population would like to increase statewide highway speed limits from 65 mph to 75 mph. How many residents does the department need to survey if it wants to be at least 99% confident that the sample proportion is within 0.05 of the true proportion?

8.4.5 The U.S. Postal Service (USPS) has used optical character recognition (OCR) since the mid-1960s. In 1983, USPS began deploying the technology to major post offices throughout the country (*www.britannica.com*). Suppose that in a random sample of 500 handwritten zip code digits, 466 were read correctly.

a. Construct a 95% confidence interval for the true proportion of correct digits that can be automatically read.

b. What sample size is needed to reduce the margin of error to 1%?

c. How would the answer to part (b) change if you had to assume that the machine read only one-half of the digits correctly?

8.4.6 Information on a packet of seeds claims that 93% of them will germinate. Of the 200 seeds that were planted, only 180 germinated.

a. Find a 95% confidence interval for the true proportion of seeds that germinate based on this sample.

b. Does this seem to provide evidence that the claim is wrong?

8.4.7 Use the data from Exercise 8.4.5 to compute the two-sided Agresti-Coull CI on the proportion of digits read correctly. Compare and discuss the relationship of this interval to the one computed in Exercise 8.4.5.

8.4.8 Use the data from Exercise 8.4.2 to compute the two-sided Agresti-Coull CI on the proportion of tears that heal. Compare and discuss the relationship of this interval to the one computed in Exercise 8.4.2.

Exercises for Section 8.6

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

8.6.1 Consider the tire-testing data described in Exercise 8.2.3. Compute a 95% prediction interval on the life of the next tire of this type tested under conditions that are similar to those employed in the original test. Compare the length of the prediction interval with the length of the 95% CI on the population mean.

8.6.2 Consider the natural frequency of beams described in Exercise 8.2.8. Compute a 90% prediction interval on the diameter of the natural frequency of the next beam of this type that will be tested. Compare the length of the prediction interval with the length of the 90% CI on the population mean.

8.6.3 Consider the television tube brightness test described in Exercise 8.2.7. Compute a 99% prediction interval on the brightness of the next tube tested. Compare the length of the prediction interval with the length of the 99% CI on the population mean.

8.6.4 Consider the test on the compressive strength of concrete described in Exercise 8.2.9. Compute a 90% prediction interval on the next specimen of concrete tested.

8.6.5 SS Consider the fuel rod enrichment data described in Exercise 8.2.11. Compute a 90% prediction interval on the enrichment of the next rod tested. Compare the length of the prediction interval with the length of the 99% CI on the population mean.

8.6.6 How would you obtain a one-sided prediction bound on a future observation? Apply this procedure to obtain a 95% one-sided prediction bound on the wall thickness of the next bottle for the situation described in Exercise 8.2.11.

8.6.7 Consider the tire-testing data in Exercise 8.2.3. Compute a 95% tolerance interval on the life of the tires that has confidence level 95%. Compare the length of the tolerance interval with the length of the 95% CI on the population mean. Which interval is shorter? Discuss the difference in interpretation of these two intervals.

8.6.8 Consider the television tube brightness data in Exercise 8.2.7. Compute a 99% tolerance interval on the brightness of the television tubes that has confidence level 95%. Compare the length of the tolerance interval with the length of the 99% CI on the population mean. Which interval is shorter? Discuss the difference in interpretation of these two intervals.

8.6.9 Consider the fuel rod enrichment data described in Exercise 8.2.11. Compute a 99% tolerance interval on rod enrichment that has confidence level 95%. Compare the length of the tolerance interval with the length of the 95% CI on the population mean.

8.6.10 Consider the strength-of-concrete data in Exercise 8.2.9. Compute a 90% tolerance interval on the compressive strength of the concrete that has 90% confidence.

Supplemental Exercises for Chapter 8

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

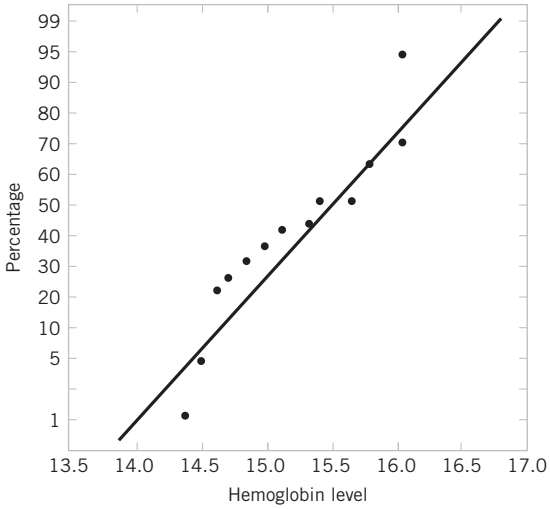
8.S7 Consider the confidence interval for μ with known standard deviation σ:

$$\bar{x} - z_{\alpha_1}\sigma/\sqrt{n} \le \mu \le \bar{x} + z_{\alpha_2}\sigma/\sqrt{n}$$

where $\alpha_1 + \alpha_2 = \alpha$. Let $\alpha = 0.05$ and find the interval for $\alpha_1 = \alpha_2 = \alpha/2 = 0.025$. Now find the interval for the case $\alpha_1 = 0.01$ and $\alpha_2 = 0.04$. Which interval is shorter? Is there any advantage to a "symmetric" confidence interval?

8.S8 WP SS An article in the *Journal of Sports Sciences* ["Iron Status in Winter Olympic Sports" (1987, Vol. 5(3), pp. 261–271)] presents the results of an investigation of the hemoglobin level of Canadian Olympic ice hockey players. The data reported are as follows (in g/dl):

15.3	16.0	14.4	16.2	16.2
14.9	15.7	15.3	14.6	15.7
16.0	15.0	15.7	16.2	14.7
14.8	14.6	15.6	14.5	15.2



a. Given the following probability plot of the data, what is a logical assumption about the underlying distribution of the data?

b. Explain why this check of the distribution underlying the sample data is important if you want to construct a confidence interval on the mean.

c. Based on this sample data, a 95% confidence interval for the mean is (15.04, 15.62). Is it reasonable to infer that the true mean could be 14.5? Explain your answer.

d. Explain why this check of the distribution underlying the sample data is important if we want to construct a confidence interval on the variance.

e. Based on these sample data, a 95% confidence interval for the variance is (0.22, 0.82). Is it reasonable to infer that the true variance could be 0.35? Explain your answer.

f. Is it reasonable to use these confidence intervals to draw an inference about the mean and variance of hemoglobin levels

(i) of Canadian doctors? Explain your answer.

(ii) of Canadian children ages 6–12? Explain your answer.

8.S9 A normal population has known mean $\mu = 50$ and variance $\sigma^2 = 5$. What is the approximate probability that the sample

variance is greater than or equal to 7.44? less than or equal to 2.56? For a random sample of size

a. $n = 16$ **b.** $n = 30$ **c.** $n = 71$

d. Compare your answers to parts (a)–(c) for the approximate probability that the sample variance is greater than or equal to 7.44. Explain why this tail probability is increasing or decreasing with increased sample size.

e. Compare your answers to parts (a)–(c) for the approximate probability that the sample variance is less than or equal to 2.56. Explain why this tail probability is increasing or decreasing with increased sample size.

8.S10 During the 1999 and 2000 baseball seasons, there was much speculation that the unusually large number of home runs hit was due at least in part to a livelier ball. One way to test the "liveliness" of a baseball is to launch the ball at a vertical surface with a known velocity V_L and measure the ratio of the outgoing velocity V_O of the ball to V_L. The ratio $R = V_O/V_L$ is called the *coefficient of restitution*. Following are measurements of the coefficient of restitution for 40 randomly selected baseballs. The balls were thrown from a pitching machine at an oak surface.

0.6248	0.6237	0.6118	0.6159	0.6298	0.6192
0.6520	0.6368	0.6220	0.6151	0.6121	0.6548
0.6226	0.6280	0.6096	0.6300	0.6107	0.6392
0.6230	0.6131	0.6223	0.6297	0.6435	0.5978
0.6351	0.6275	0.6261	0.6262	0.6262	0.6314
0.6128	0.6403	0.6521	0.6049	0.6170	
0.6134	0.6310	0.6065	0.6214	0.6141	

a. Is there evidence to support the assumption that the coefficient of restitution is normally distributed?

b. Find a 99% CI on the mean coefficient of restitution.

c. Find a 99% prediction interval on the coefficient of restitution for the next baseball that will be tested.

d. Find an interval that will contain 99% of the values of the coefficient of restitution with 95% confidence.

e. Explain the difference in the three intervals computed in parts (b), (c), and (d).

8.S11 WP Consider the baseball coefficient of restitution data in Exercise 8.S10. Suppose that any baseball that has a coefficient of restitution that exceeds 0.635 is considered too lively. Based on the available data, what proportion of the baseballs in the sampled population are too lively? Find a 95% lower confidence bound on this proportion.

8.S12 WP An article in the *ASCE Journal of Energy Engineering* ["Overview of Reservoir Release Improvements at 20 TVA Dams" (Vol. 125, April 1999, pp.s 1–17)] presents data on dissolved oxygen concentrations in streams below 20 dams in the Tennessee Valley Authority system. The observations are (in milligrams per liter): 5.0, 3.4, 3.9, 1.3, 0.2, 0.9, 2.7, 3.7, 3.8, 4.1, 1.0, 1.0, 0.8, 0.4, 3.8, 4.5, 5.3, 6.1, 6.9, and 6.5.

a. Is there evidence to support the assumption that the dissolved oxygen concentration is normally distributed?

b. Find a 95% CI on the mean dissolved oxygen concentration.

c. Find a 95% prediction interval on the dissolved oxygen concentration for the next stream in the system that will be tested.

d. Find an interval that will contain 95% of the values of the dissolved oxygen concentration with 99% confidence.

e. Explain the difference in the three intervals computed in parts (b), (c), and (d).

8.S13 WP An article in *The Engineer* ("Redesign for Suspect Wiring," June 1990) reported the results of an investigation into wiring errors on commercial transport aircraft that may display faulty information to the flight crew. Such a wiring error may have been responsible for the crash of a British Midland Airways aircraft in January 1989 by causing the pilot to shut down the wrong engine. Of 1600 randomly selected aircraft, 8 were found to have wiring errors that could display incorrect information to the flight crew.

a. Find a 99% confidence interval on the proportion of aircraft that have such wiring errors.

b. Suppose that you use the information in this example to provide a preliminary estimate of p. How large a sample would be required to produce an estimate of p that we are 99% confident differs from the true value by at most 0.008?

c. Suppose that you did not have a preliminary estimate of p. How large a sample would be required if you wanted to be at least 99% confident that the sample proportion differs from the true proportion by at most 0.008 regardless of the true value of p?

d. Comment on the usefulness of preliminary information in computing the needed sample size.

8.S14 An article in the *Journal of Human Nutrition and Dietetics* ["The Validation of Energy and Protein Intakes by Doubly Labeled Water and 24-Hour Urinary Nitrogen Excretion in Post-Obese Subjects" (1995, Vol. 8, pp. 51–64)] showed the energy intake expressed as a basal metabolic rate, BMR (MJ).

5.40	5.67	5.79	6.85	6.92
5.70	6.08	5.48	5.44	5.51

a. Use a normal probability plot to check the normality assumption.

b. Find a 99% two-sided confidence interval on the mean BMR.

Chapter 9 Exercises

Exercises for Section 9.1

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

9.1.1 WP State whether each of the following situations is a correctly stated hypothesis testing problem and why.

a. H_0: $\mu = 25$, H_1: $\mu \neq 25$ **b.** H_0: $\sigma > 10$, H_1: $\sigma = 10$
c. H_0: $\bar{x} = 50$, H_1: $\bar{x} \neq 50$ **d.** H_0: $p = 0.1$, H_1: $p = 0.5$
e. H_0: $s = 30$, H_1: $s > 30$

9.1.2 A semiconductor manufacturer collects data from a new tool and conducts a hypothesis test with the null hypothesis that a critical dimension mean width equals 100 nm. The conclusion is to not reject the null hypothesis. Does this result provide strong evidence that the critical dimension mean equals 100 nm? Explain.

9.1.3 WP SS A textile fiber manufacturer is investigating a new drapery yarn, which the company claims has a mean thread elongation of 12 kilograms with a standard deviation of 0.5 kilograms. The company wishes to test the hypothesis H_0: $\mu = 12$ against H_1: $\mu < 12$, using a random sample of four specimens.

a. What is the type I error probability if the critical region is defined as $\bar{x} < 11.5$ kilograms?

b. Find β for the case in which the true mean elongation is 11.25 kilograms.

c. Find β for the case in which the true mean is 11.5 kilograms.

9.1.4 Repeat Exercise 9.1.3 using a sample size of $n = 16$ and the same critical region.

9.1.5 WP In Exercise 9.1.3, calculate the P-value if the observed statistic is

a. $\bar{x} = 11.25$ **b.** $\bar{x} = 11.0$ **c.** $\bar{x} = 11.75$

9.1.6 In Exercise 9.1.3, calculate the probability of a type II error if the true mean elongation is 11.5 kilograms and

a. $\alpha = 0.05$ and $n = 4$ **b.** $\alpha = 0.05$ and $n = 16$

c. Compare the values of β calculated in the previous parts. What conclusion can you draw?

9.1.7 WP VS SS A consumer products company is formulating a new shampoo and is interested in foam height (in millimeters). Foam height is approximately normally distributed and has a standard deviation of 20 millimeters. The company wishes to test H_0: $\mu = 175$ millimeters versus H_1: $\mu > 175$ millimeters, using the results of $n = 10$ samples.

a. Find the type I error probability α if the critical region is $\bar{x} > 185$.

b. What is the probability of type II error if the true mean foam height is 185 millimeters?

c. Find β for the true mean of 195 millimeters.

9.1.8 In Exercise 9.1.7, calculate the probability of a type II error if the true mean foam height is 185 millimeters and

a. $\alpha = 0.05$ and $n = 10$ **b.** $\alpha = 0.05$ and $n = 16$

c. Compare the values of β calculated in the previous parts. What conclusion can you draw?

9.1.9 WP In Exercise 9.1.7, find the boundary of the critical region if the type I error probability is

a. $\alpha = 0.01$ and $n = 10$ **b.** $\alpha = 0.05$ and $n = 10$
c. $\alpha = 0.01$ and $n = 16$ **d.** $\alpha = 0.05$ and $n = 16$

9.1.10 WP A manufacturer is interested in the output voltage of a power supply used in a PC. Output voltage is assumed to be normally distributed with standard deviation 0.25 volt, and the manufacturer wishes to test H_0: $\mu = 5$ volts against H_1: $\mu \neq 5$ volts, using $n = 8$ units.

a. The acceptance region is $4.85 \leq \bar{x} \leq 5.15$. Find the value of α.

b. Find the power of the test for detecting a true mean output voltage of 5.1 volts.

9.1.11 The proportion of adults living in Tempe, Arizona, who are college graduates is estimated to be $p = 0.4$. To test this hypothesis, a random sample of 15 Tempe adults is selected. If the number of college graduates is between 4 and 8, the hypothesis will be accepted; otherwise, you will conclude that $p \neq 0.4$.

a. Find the type I error probability for this procedure, assuming that $p = 0.4$.

b. Find the probability of committing a type II error if the true proportion is really $p = 0.2$.

9.1.12 The proportion of residents in Phoenix favoring the building of toll roads to complete the freeway system is believed to be $p = 0.3$. If a random sample of 10 residents shows that one or fewer favor this proposal, we will conclude that $p < 0.3$.

a. Find the probability of type I error if the true proportion is $p = 0.3$.

b. Find the probability of committing a type II error with this procedure if $p = 0.2$.

c. What is the power of this procedure if the true proportion is $p = 0.2$?

9.1.13 If we plot the probability of accepting H_0: $\mu = \mu_0$ versus various values of μ and connect the points with a smooth curve, we obtain the **operating characteristic curve** (or the **OC curve**) of the test procedure. These curves are used extensively in industrial applications of hypothesis testing to display the sensitivity and relative performance of the test. When the true mean is really equal to μ_0, the probability of accepting H_0 is $1 - \alpha$.

a. Construct an OC curve for Exercise 9.1.7, using values of the true mean μ of 178, 181, 184, 187, 190, 193, 196, and 199.

b. Convert the OC curve into a plot of the **power function** of the test.

Exercises for Section 9.2

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

9.2.1 State the null and alternative hypothesis in each case.

a. A hypothesis test will be used to potentially provide evidence that the population mean is more than 10.

b. A hypothesis test will be used to potentially provide evidence that the population mean is not equal to 7.

c. A hypothesis test will be used to potentially provide evidence that the population mean is less than 5.

9.2.2 WP A hypothesis will be used to test that a population mean equals 5 against the alternative that the population mean is less than 5 with known variance σ. What is the critical value for the test statistic Z_0 for the following significance levels?

a. 0.01 **b.** 0.05 **c.** 0.10

9.2.3 Output from a software package follows:

One-Sample Z:

```
Test of mu = 20 vs > 20
The assumed standard deviation = 0.75
Variable   N    Mean    StDev   SE Mean   Z    P
   x      10  19.889     ?      0.237    ?    ?
```

a. Fill in the missing items. What conclusions would you draw?

b. Is this a one-sided or a two-sided test?

c. Use the normal table and the preceding data to construct a 95% two-sided CI on the mean.

d. What would the P-value be if the alternative hypothesis is $H_1: \mu \neq 20$?

9.2.4 WP For the hypothesis test $H_0: \mu = 10$ against $H_1: \mu > 10$ and variance known, calculate the P-value for each of the following test statistics.

a. $z_0 = 2.05$ **b.** $z_0 = -1.84$ **c.** $z_0 = 0.4$

9.2.5 SS Output from a software package follows:

One-Sample Z:

```
Test of mu = 99 vs > 99
The assumed standard deviation = 2.5
Variable  N    Mean   StDev  SE Mean   Z     P
   x     12 100.039  2.365     ?     1.44  0.075
```

a. Fill in the missing items. What conclusions would you draw?

b. Is this a one-sided or a two-sided test?

c. If the hypothesis had been $H_0: \mu = 98$ versus $H_0: \mu > 98$, would you reject the null hypothesis at the 0.05 level of significance? Can you answer this without referring to the normal table?

d. Use the normal table and the preceding data to construct a 95% lower bound on the mean.

e. What would the P-value be if the alternative hypothesis is $H_1: \mu \neq 99$?

9.2.6 A manufacturer produces crankshafts for an automobile engine. The crankshafts wear after 100,000 miles (0.0001 inch) is of interest because it is likely to have an impact on warranty claims. A random sample of $n = 15$ shafts is tested and $\bar{x} = 2.78$. It is known that $\sigma = 0.9$ and that wear is normally distributed.

a. Test $H_0: \mu = 3$ versus $H_1: \mu \neq 3$ using $\alpha = 0.05$.

b. What is the power of this test if $\mu = 3.25$?

c. What sample size would be required to detect a true mean of 3.75 if we wanted the power to be at least 0.9?

9.2.7 WP The life in hours of a battery is known to be approximately normally distributed with standard deviation $\sigma = 1.25$ hours. A random sample of 10 batteries has a mean life of $\bar{x} = 40.5$ hours.

a. Is there evidence to support the claim that battery life exceeds 40 hours? Use $\alpha = 0.5$.

b. What is the P-value for the test in part (a)?

c. What is the β-error for the test in part (a) if the true mean life is 42 hours?

d. What sample size would be required to ensure that β does not exceed 0.10 if the true mean life is 44 hours?

e. Explain how you could answer the question in part (a) by calculating an appropriate confidence bound on battery life.

9.2.8 WP *Supercavitation* is a propulsion technology for undersea vehicles that can greatly increase their speed. It occurs above approximately 50 meters per second when pressure drops sufficiently to allow the water to dissociate into water vapor, forming a gas bubble behind the vehicle. When the gas bubble completely encloses the vehicle, supercavitation is said to occur. Eight tests were conducted on a scale model of an undersea vehicle in a towing basin with the average observed speed $\bar{x} = 102.2$ meters per second. Assume that speed is normally distributed with known standard deviation $\sigma = 4$ meters per second.

a. Test the hypothesis $H_0: \mu = 100$ versus $H_1: \mu < 100$ using $\alpha = 0.05$.

b. What is the P-value for the test in part (a)?

c. Compute the power of the test if the true mean speed is as low as 95 meters per second.

d. What sample size would be required to detect a true mean speed as low as 95 meters per second if you wanted the power of the test to be at least 0.85?

e. Explain how the question in part (a) could be answered by constructing a one-sided confidence bound on the mean speed.

9.2.9 WP VS SS Medical researchers have developed a new artificial heart constructed primarily of titanium and plastic. The heart will last and operate almost indefinitely once it is implanted in the patient's body, but the battery pack needs to be recharged

about every 4 hours. A random sample of 50 battery packs is selected and subjected to a life test. The average life of these batteries is 4.05 hours. Assume that battery life is normally distributed with standard deviation $\sigma = 0.2$ hour.

a. Is there evidence to support the claim that mean battery life exceeds 4 hours? Use $\alpha = 0.05$.

b. What is the P-value for the test in part (a)?

c. Compute the power of the test if the true mean battery life is 4.5 hours.

d. What sample size would be required to detect a true mean battery life of 4.5 hours if you wanted the power of the test to be at least 0.9?

e. Explain how the question in part (a) could be answered by constructing a one-sided confidence bound on the mean life.

9.2.10 The bacterial strain *Acinetobacter* has been tested for its adhesion properties. A sample of five measurements gave readings of 2.69, 5.76, 2.67, 1.62 and 4.12 dyne-cm^2. Assume that the standard deviation is known to be 0.66 dyne-cm^2 and that the scientists are interested in high adhesion (at least 2.5 dyne-cm^2).

a. Should the alternative hypothesis be one-sided or two-sided?

b. Test the hypothesis that the mean adhesion is 2.5 dyne-cm^2.

c. What is the P-value of the test statistic?

Exercises for Section 9.3

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

9.3.1 A hypothesis will be used to test that a population mean equals 10 against the alternative that the population mean is greater than 10 with unknown variance. What is the critical value for the test statistic T_0 for the following significance levels?

a. $\alpha = 0.01$ and $n = 20$ **b.** $\alpha = 0.05$ and $n = 12$

c. $\alpha = 0.10$ and $n = 15$

9.3.2 Consider the following computer output.

One-Sample T:

```
Test of mu = 91 vs > 91

                                          95%
                                        Lower
Variable  N    Mean   StDev  SE Mean   Bound  T  P
    x     20 92.379  0.717      ?       ?    ?  ?
```

a. Fill in the missing values. You may calculate bounds on the P-value. What conclusions would you draw?

b. Is this a one-sided or a two-sided test?

c. If the hypothesis had been H_0: $\mu = 90$ versus H_1: $\mu > 90$, would your conclusions change?

9.3.3 SS For the hypothesis test H_0: $\mu = 7$ against H_1: $\mu \neq 7$ with variance unknown and $n = 20$, approximate the P-value for each of the following test statistics.

a. $t_0 = 2.05$ **b.** $t_0 = -1.84$ **c.** $t_0 = 0.4$

9.3.4 Consider the following computer output.

One-Sample T:

```
Test of mu = 34 vs not = 34

Variable N   Mean  StDev SE Mean      95% CI      T   P
    x    16 35.274 1.783    ?    (34.324, 36.224) ? 0.012
```

a. How many degrees of freedom are there on the t-test statistic?

b. Fill in the missing quantities.

c. At what level of significance can the null hypothesis be rejected?

d. If the hypothesis had been H_0: $\mu = 34$ versus H_1: $\mu > 34$, would the P-value have been larger or smaller?

e. If the hypothesis had been H_0: $\mu = 34.5$ versus H_1: $\mu \neq 34.5$, would you have rejected the null hypothesis at the 0.05 level?

9.3.5 WP For the hypothesis test H_0: $\mu = 5$ against H_1: $\mu < 5$ with variance unknown and $n = 12$, approximate the P-value for each of the following test statistics.

a. $t_0 = 2.05$ **b.** $t_0 = -1.84$ **c.** $t_0 = 0.4$

9.3.6 An article in the *ASCE Journal of Energy Engineering* (1999, Vol. 125, pp. 59–75) describes a study of the thermal inertia properties of autoclaved aerated concrete used as a building material. Five samples of the material were tested in a structure, and the average interior temperatures (°C) reported were as follows: 23.01, 22.22, 22.04, 22.62, and 22.59.

a. Test the hypotheses H_0: $\mu = 22.5$ versus H_1: $\mu \neq 22.5$, using $\alpha = 0.05$. Find the P-value.

b. Check the assumption that interior temperature is normally distributed.

c. Compute the power of the test if the true mean interior temperature is as high as 22.75.

d. What sample size would be required to detect a true mean interior temperature as high as 22.75 if you wanted the power of the test to be at least 0.9?

e. Explain how the question in part (a) could be answered by constructing a two-sided confidence interval on the mean interior temperature.

9.3.7 VS An article in *Growth: A Journal Devoted to Problems of Normal and Abnormal Growth* ["Comparison of Measured and Estimated Fat-Free Weight, Fat, Potassium and Nitrogen of Growing Guinea Pigs" (1982, Vol. 46(4), pp. 306–321)] reported the results of a study that measured the body weight (in grams) for guinea pigs at birth.

421.0	452.6	456.1	494.6	373.8
90.5	110.7	96.4	81.7	102.4
241.0	296.0	317.0	290.9	256.5
447.8	687.6	705.7	879.0	88.8
296.0	273.0	268.0	227.5	279.3
258.5	296.0			

a. Test the hypothesis that mean body weight is 300 grams. Use $\alpha = 0.05$.

b. What is the smallest level of significance at which you would be willing to reject the null hypothesis?

c. Explain how you could answer the question in part (a) with a two-sided confidence interval on mean body weight.

9.3.8 WP Cloud seeding has been studied for many decades as a weather modification procedure (for an interesting study of this subject, see the article in *Technometrics*, "A Bayesian Analysis of a Multiplicative Treatment Effect in Weather Modification," 1975, Vol. 17, pp. 161–166). The rainfall in acre-feet from 20 clouds that were selected at random and seeded with silver nitrate follows: 18.0, 30.7, 19.8, 27.1, 22.3, 18.8, 31.8, 23.4, 21.2, 27.9, 31.9, 27.1, 25.0, 24.7, 26.9, 21.8, 29.2, 34.8, 26.7, and 31.6.

a. Can you support a claim that mean rainfall from seeded clouds exceeds 25 acre-feet? Use $\alpha = 0.01$. Find the P-value.

b. Check that rainfall is normally distributed.

c. Compute the power of the test if the true mean rainfall is 27 acre-feet.

d. What sample size would be required to detect a true mean rainfall of 27.5 acre-feet if you wanted the power of the test to be at least 0.9?

e. Explain how the question in part (a) could be answered by constructing a one-sided confidence bound on the mean diameter.

9.3.9 WP SS A 1992 article in the *Journal of the American Medical Association* ("A Critical Appraisal of 98.6 Degrees F, the Upper Limit of the Normal Body Temperature, and Other Legacies of Carl Reinhold August Wunderlich") reported body temperature, gender, and heart rate for a number of subjects. The body temperatures for 25 female subjects follow: 97.8, 97.2, 97.4, 97.6, 97.8, 97.9, 98.0, 98.0, 98.0, 98.1, 98.2, 98.3, 98.3, 98.4, 98.4, 98.4, 98.5, 98.6, 98.6, 98.7, 98.8, 98.8, 98.9, 98.9, and 99.0.

a. Test the hypothesis H_0: $\mu = 98.6$ versus H_1: $\mu \neq 98.6$, using $\alpha = 0.05$. Find the P-value.

b. Check the assumption that female body temperature is normally distributed.

c. Compute the power of the test if the true mean female body temperature is as low as 98.0.

d. What sample size would be required to detect a true mean female body temperature as low as 98.2 if you wanted the power of the test to be at least 0.9?

e. Explain how the question in part (a) could be answered by constructing a two-sided confidence interval on the mean female body temperature.

9.3.10 WP Reconsider the data from *Medicine and Science in Sports and Exercise* described in Exercise 8.2.6. The sample size was seven and the sample mean and sample standard deviation were 315 watts and 16 watts, respectively.

a. Is there evidence that leg strength exceeds 300 watts at significance level 0.05? Find the P-value.

b. Compute the power of the test if the true strength is 305 watts.

c. What sample size would be required to detect a true mean of 305 watts if the power of the test should be at least 0.90?

d. Explain how the question in part (a) could be answered with a confidence interval.

9.3.11 Exercise 6.2.8 gave data on the heights of female engineering students at ASU.

a. Can you support a claim that the mean height of female engineering students at ASU is at least 65 inches? Use $\alpha = 0.05$. Find the P-value.

b. Check the normality assumption.

c. Compute the power of the test if the true mean height is 68 inches.

d. What sample size would be required to detect a true mean height of 66 inches if you wanted the power of the test to be at least 0.8?

9.3.12 Exercise 6.2.10 presented data on the concentration of suspended solids in lake water.

a. Test the hypothesis H_0: $\mu = 55$ versus H_1: $\mu \neq 55$; use $\alpha = 0.05$. Find the P-value.

b. Check the normality assumption.

c. Compute the power of the test if the true mean concentration is as low as 50.

d. What sample size would be required to detect a true mean concentration as low as 50 if you wanted the power of the test to be at least 0.9?

9.3.13 In a little over a month, from June 5, 1879, to July 2, 1879, Albert Michelson measured the velocity of light in air 100 times (Stigler, *Annals of Statistics*, 1977). Today we know that the true value is 299,734.5 km/sec. Michelson's data have a mean of 299,852.4 km/sec with a standard deviation of 79.01.

a. Find a two-sided 95% confidence interval for the true mean (the true value of the speed of light).

b. What does the confidence interval say about the accuracy of Michelson's measurements?

Exercises for Section 9.4

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

9.4.1 WP Consider the test of H_0: $\sigma^2 = 7$ against H_1: $\sigma^2 \neq 7$. What are the critical values for the test statistic χ_0^2 for the following significance levels and sample sizes?

a. $\alpha = 0.01$ and $n = 20$ **b.** $\alpha = 0.05$ and $n = 12$
c. $\alpha = 0.10$ and $n = 15$

9.4.2 WP Consider the test of H_0: $\sigma^2 = 5$ against H_1: $\sigma^2 < 5$. Approximate the P-value for each of the following test statistics.

a. $x_0^2 = 25.2$ and $n = 20$ **b.** $x_0^2 = 15.2$ and $n = 12$
c. $x_0^2 = 4.2$ and $n = 15$

9.4.3 The data from *Medicine and Science in Sports and Exercise* described in Exercise 8.3.4 considered ice hockey player performance after electrostimulation training. In summary, there were 17 players, and the sample standard deviation of performance was 0.09 seconds.

a. Is there strong evidence to conclude that the standard deviation of performance time exceeds the historical value of 0.75 seconds? Use $\alpha = 0.05$. Find the P-value for this test.
b. Discuss how part (a) could be answered by constructing a 95% one-sided confidence interval for σ.

9.4.4 WP Reconsider the percentage of titanium in an alloy used in aerospace castings from Exercise 8.3.1. Recall that $s = 0.37$ and $n = 51$.

a. Test the hypothesis H_0: $\sigma = 0.25$ versus H_1: $\sigma \neq 0.25$ using $\alpha = 0.05$. State any necessary assumptions about the underlying distribution of the data. Find the P-value.
b. Explain how you could answer the question in part (a) by constructing a 95% two-sided confidence interval for σ.

9.4.5 WP SS Data for tire life was described in Exercise 8.2.3. The sample standard deviation was 3645.94 kilometers and $n = 16$.

a. Can you conclude, using $\alpha = 0.05$, that the standard deviation of tire life is less than 4000 kilometers? State any necessary assumptions about the underlying distribution of the data. Find the P-value for this test.
b. Explain how you could answer the question in part (a) by constructing a 95% one-sided confidence interval for σ.

Exercises for Section 9.5

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

9.5.1 Consider the following computer output.

Test and CI for One Proportion

Test of p = 0.4 vs p not = 0.4

X	N	Sample p	95% CI	Z-Value	P-Value
98	275	?	(0.299759, 0.412968)	?	?

Using the normal approximation,

a. Is this a one-sided or a two-sided test?
b. Complete the missing items.
c. The normal approximation was used in the problem. Was that appropriate?

9.5.2 WP Suppose that of 1000 customers surveyed, 850 are satisfied or very satisfied with a corporation's products and services.

a. Test the hypothesis H_0: $p = 0.9$ against H_1: $p \neq 0.9$ at $\alpha = 0.05$. Find the P-value.
b. Explain how the question in part (a) could be answered by constructing a 95% two-sided confidence interval for p.

9.5.3 An article in the *British Medical Journal* ["Comparison of Treatment of Renal Calculi by Operative Surgery, Percutaneous Nephrolithotomy, and Extra-Corporeal Shock Wave Lithotripsy" (1986, Vol. 292, pp. 879–882)] repeated that percutaneous nephrolithotomy (PN) had a success rate in removing kidney stones of 289 of 350 patients. The traditional method was 78% effective.

a. Is there evidence that the success rate for PN is greater than the historical success rate? Find the P-value.
b. Explain how the question in part (a) could be answered with a confidence interval.

9.5.4 WP An article in *Fortune* (September 21, 1992) claimed that nearly one-half of all engineers continue academic studies beyond the B.S. degree, ultimately receiving either an M.S. or a Ph.D. degree. Data from an article in *Engineering Horizons* (Spring 1990) indicated that 117 of 484 new engineering graduates were planning graduate study.

a. Are the data from *Engineering Horizons* consistent with the claim reported by *Fortune*? Use $\alpha = 0.05$ in reaching your conclusions. Find the P-value for this test.
b. Discuss how you could have answered the question in part (a) by constructing a two-sided confidence interval on p.

9.5.5 WP SS A random sample of 500 registered voters in Phoenix is asked if they favor the use of oxygenated fuels year-round to reduce air pollution. If more than 315 voters respond positively, we will conclude that at least 60% of the voters favor the use of these fuels.

a. Find the probability of type I error if exactly 60% of the voters favor the use of these fuels.

b. What is the type II error probability β if 75% of the voters favor this action?

9.5.6 A computer manufacturer ships laptop computers with the batteries fully charged so that customers can begin to use their purchases right out of the box. In its last model, 85% of customers received fully charged batteries. To simulate arrivals, the company shipped 100 new model laptops to various company sites around the country. Of the 100 laptops shipped, 96 of them arrived reading 100% charged. Do the data provide evidence that this model's rate is at least as high as the previous model? Test the hypothesis at $\alpha = 0.05$.

Exercises for Section 9.7

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

9.7.1 WP Consider the following frequency table of observations on the random variable X.

Values	0	1	2	3	4
Frequency	4	21	10	13	2

a. Based on these 50 observations, is a binomial distribution with $n = 6$ and $p = 0.25$ an appropriate model? Perform a goodness-of-fit procedure with $\alpha = 0.05$.

b. Calculate the P-value for this test.

9.7.2 WP Consider the following frequency table of observations on the random variable X.

Values	0	1	2	3	4
Observed frequency	24	30	31	11	4

a. Based on these 100 observations, is a Poisson distribution with a mean of 1.2 an appropriate model? Perform a goodness-of-fit procedure with $\alpha = 0.05$.

b. Calculate the P-value for this test.

9.7.3 SS A group of civil engineering students has tabulated the number of cars passing eastbound through the intersection of Mill and University Avenues. They obtained the data in the following table.

Vehicles per Minute	Observed Frequency	Vehicles per Minute	Observed Frequency
40	14	53	102
41	24	54	96
42	57	55	90
43	111	56	81
44	194	57	73
45	256	58	64
46	296	59	61
47	378	60	59
48	250	61	50
49	185	62	42
50	171	63	29
51	150	64	18
52	110	65	15

a. Does the assumption of a Poisson distribution seem appropriate as a probability model for this process? Use $\alpha = 0.05$.

b. Calculate the P-value for this test.

9.7.4 WP The number of calls arriving at a switchboard from noon to 1:00 P.M. during the business days Monday through Friday is monitored for 6 weeks (i.e., 30 days). Let X be defined as the number of calls during that one-hour period. The relative frequency of calls was recorded and reported as

Value	5	6	8	9	10
Relative frequency	0.067	0.067	0.100	0.133	0.200
Value	11	12	13	14	15
Relative frequency	0.133	0.133	0.067	0.033	0.067

a. Does the assumption of a Poisson distribution seem appropriate as a probability model for this data? Use $\alpha = 0.05$.

b. Calculate the P-value for this test.

Exercises for Section 9.8

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

9.8.1 The Hopkins Forest is a 2600-acre forest reserve located at the intersection of three states: New York, Vermont, and Massachusetts. Researchers monitor forest resources to study long-term ecological changes. They have conducted surveys of existing trees, shrubs, and herbs at various sites in the forest for nearly 100 years. Following are some data from surveys of three species of maple trees at the same location over three very different time periods.

		Species (Acre)			
		Pennsylvanicum	Rubrum	Saccharum	Total
Year	1936	12	27	94	133
	1972	22	40	52	114
	2011	97	25	18	140
	Total	131	92	164	387

Does the species distribution seem to be independent of year? Test the hypothesis at $\alpha = 0.05$. Find the *P*-value of the test statistic.

9.8.2 Did survival rate for passengers on the *Titanic* really depend on the type of ticket they had? Following are the data for the 2201 people on board listed by whether they survived and what type of ticket they had. Does survival appear to be independent of ticket class? (Test the hypothesis at $\alpha = 0.05$.) What is the *P*-value of the test statistic?

	Crew	First	Second	Third	Total
Alive	212	202	118	178	710
Dead	673	123	167	528	1491
Total	885	325	285	706	2201

9.8.3 WP A company operates four machines in three shifts each day. From production records, the following data on the number of breakdowns are collected:

	Machines			
Shift	A	B	C	D
1	41	20	12	16
2	31	11	9	14
3	15	17	16	10

Test the hypothesis (using $\alpha = 0.05$) that breakdowns are independent of the shift. Find the *P*-value for this test.

9.8.4 WP Patients in a hospital are classified as surgical or medical. A record is kept of the number of times patients require nursing service during the night and whether or not these patients are on Medicare. The data are presented here:

	Patient Category	
Medicare	Surgical	Medical
Yes	46	52
No	36	43

Test the hypothesis (using $\alpha = 0.01$) that calls by surgical-medical patients are independent of whether the patients are receiving Medicare. Find the *P*-value for this test.

9.8.5 WP SS Grades in a statistics course and an operations research course taken simultaneously were as follows for a group of students.

	Operation Research Grade			
Statistics Grade	A	B	C	Other
A	25	6	17	13
B	17	16	15	6
C	18	4	18	10
Other	10	8	11	20

Are the grades in statistics and operations research related? Use $\alpha = 0.01$ in reaching your conclusion. What is the *P*-value for this test?

9.8.6 WP An experiment with artillery shells yields the following data on the characteristics of lateral deflections and ranges. Would you conclude that deflection and range are independent? Use $\alpha = 0.05$. What is the *P*-value for this test?

	Lateral Deflection		
Range (yards)	Left	Normal	Right
0–1,999	6	14	8
2,000–5,999	9	11	4
6,000–11,999	8	17	6

9.8.7 WP A study is being made of the failures of an electronic component. There are four types of failures possible and two mounting positions for the device. The following data have been taken:

	Failure Type			
Mounting Position	A	B	C	D
1	22	46	18	9
2	4	17	6	12

Would you conclude that the type of failure is independent of the mounting position? Use $\alpha = 0.01$. Find the *P*-value for this test.

9.8.8 WP A random sample of students is asked their opinions on a proposed core curriculum change. The results are as follows.

	Opinion	
Class	**Favoring**	**Opposing**
Freshman	120	80
Sophomore	70	130
Junior	60	70
Senior	40	60

Test the hypothesis that opinion on the change is independent of class standing. Use $\alpha = 0.05$. What is the P-value for this test?

9.8.9 WP An article in the *British Medical Journal* ["Comparison of Treatment of Renal Calculi by Operative Surgery, Percutaneous Nephrolithotomy, and Extracorporeal Shock Wave Lithotripsy" (1986, Vol. 292, pp. 879–882)] reported that percutaneous nephrolithotomy (PN) had a success rate in removing kidney stones of 289 out of 350 (83%) patients. However, when the stone diameter was considered, the results looked different. For stones of <2 cm, 87% (234/270) of cases were successful. For stones of ≥2 cm, a success rate of 69% (55/80) was observed for PN.

a. Are the successes and size of stones independent? Use $\alpha = 0.05$.

b. Find the P-value for this test.

Exercises for Section 9.9

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

9.9.1 WP The titanium content in an aircraft-grade alloy is an important determinant of strength. A sample of 20 test coupons reveals the following titanium content (in percent):

8.32, 8.05, 8.93, 8.65, 8.25, 8.46, 8.52, 8.35, 8.36, 8.41, 8.42, 8.30, 8.71, 8.75, 8.60, 8.83, 8.50, 8.38, 8.29, 8.46

The median titanium content should be 8.5%.

a. Use the sign test with $\alpha = 0.05$ to investigate this hypothesis. Find the P-value for this test.

b. Use the normal approximation for the sign test to test H_0: $\tilde{\mu} = 8.5$ versus H_1: $\tilde{\mu} \neq 8.5$ with $\alpha = 0.05$. What is the P-value for this test?

9.9.2 WP Ten samples were taken from a plating bath used in an electronics manufacturing process, and the pH of the bath was determined. The sample pH values are 7.91, 7.85, 6.82, 8.01, 7.46, 6.95, 7.05, 7.35, 7.25, and 7.42. Manufacturing engineering believes that pH has a median value of 7.0.

a. Do the sample data indicate that this statement is correct? Use the sign test with $\alpha = 0.05$ to investigate this hypothesis. Find the P-value for this test.

b. Use the normal approximation for the sign test to test H_0: $\tilde{\mu} = 7.0$ versus H_0: $\tilde{\mu} \neq 7.0$. What is the P-value for this test?

9.9.3 WP An inspector measured the diameter of a ball bearing using a new type of caliper. The results were as follows (in mm): 0.265, 0.263, 0.266, 0.267, 0.267, 0.265, 0.267,0.267, 0.265, 0.268, 0.268, and 0.263.

a. Use the Wilcoxon signed-rank test to evaluate the claim that the mean ball diameter is 0.265 mm. Use $\alpha = 0.05$.

b. Use the normal approximation for the test. With $\alpha = 0.05$, what conclusions can you draw?

9.9.4 WP The impurity level (in ppm) is routinely measured in an intermediate chemical product. The following data were observed in a recent test:

2.4, 2.5, 1.7, 1.6, 1.9, 2.6, 1.3, 1.9, 2.0, 2.5, 2.6, 2.3, 2.0, 1.8, 1.3, 1.7, 2.0, 1.9, 2.3, 1.9, 2.4, 1.6

Can you claim that the median impurity level is less than 2.5 ppm?

a. State and test the appropriate hypothesis using the sign test with $\alpha = 0.05$. What is the P-value for this test?

b. Use the normal approximation for the sign test to test H_0: $\tilde{\mu} = 2.5$ versus H_1: $\tilde{\mu} < 2.5$. What is the P-value for this test?

9.9.5 WP SS A primer paint can be used on aluminum panels. The primer's drying time is an important consideration in the manufacturing process. Twenty panels are selected, and the drying times are as follows: 1.6, 1.3, 1.5, 1.6, 1.7, 1.9, 1.8, 1.6, 1.4, 1.8, 1.9, 1.8, 1.7, 1.5, 1.6, 1.4, 1.3, 1.6, 1.5, and 1.8. Is there evidence that the mean drying time of the primer exceeds 1.5 hr?

9.9.6 WP A new type of tip can be used in a Rockwell hardness tester. Eight coupons from test ingots of a nickel-based alloy are selected, and each coupon is tested using the new tip. The Rockwell C-scale hardness readings are 63, 65, 58, 60, 55, 57, 53, and 59. Do the results support the claim that the mean hardness exceeds 60 at a 0.05 level?

Exercises for Section 9.10

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

9.10.1 A chemical products manufacturer must identify a new supplier for a raw material that is an essential component of a particular product. The previous supplier was able to deliver material with a mean molecular weight of 3500. The new supplier must show equivalence to this value of molecular weight. If the new supplier can deliver material that has a molecular weigh that is within 50 units of this value, it will be considered equivalent. A random sample of 10 lots of product is available, and the sample mean and standard deviation of molecular weight are 3550 and 25, respectively.

a. State the appropriate hypotheses that must be tested to demonstrate equivalence.

b. What are your conclusions using $\alpha = 0.05$?

9.10.2 In developing a generic drug, it is necessary for a manufacturer of biopharmaceutical products to show equivalence to the current product. The variable of interest is the absorption rate of the product. The current product has an absorption rate of 18 mg/hr. If the new generic product has an absorption rate that is within 0.50 mg/hr of this value, it will be considered equivalent. A random sample of 20 units of product is available, and the sample mean and standard deviation of absorption rate are 18.22 mg/hr and 0.92 mg/hr, respectively.

a. State the appropriate hypotheses that must be tested to demonstrate equivalence.

b. What are your conclusions using $\alpha = 0.05$?

9.10.3 SS The mean bond strength of a cement product must be at least 10000 psi. The process by which this material is manufactured must show equivalence to this standard. If the process can manufacture cement for which the mean bond strength is at least 9750 psi, it will be considered equivalent to the standard. A random sample of six observations is available, and the sample mean and standard deviation of bond strength are 9360 psi and 42.6 psi, respectively.

a. State the appropriate hypotheses that must be tested to demonstrate equivalence.

b. What are your conclusions using $\alpha = 0.05$?

9.10.4 The mean breaking strength of a ceramic insulator must be at least 10 psi. The process by which this insulator is manufactured must show equivalence to this standard. If the process can manufacture insulators with a mean breaking strength of at least 9.5 psi, it will be considered equivalent to the standard. A random sample of 50 insulators is available, and the sample mean and standard deviation of breaking strength are 9.31 psi and 0.22 psi, respectively.

a. State the appropriate hypotheses that must be tested to demonstrate equivalence.

b. What are your conclusions using $\alpha = 0.05$?

Exercises for Section 9.11

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

9.11.1 Suppose that eight sets of hypotheses about a population proportion of the form

$$H_0: p = 0.3 \qquad H_1: p > 0.3$$

have been tested and that the P-values for these tests are 0.15, 0.83, 0.103, 0.024, 0.03, 0.07, 0.09, and 0.13. Use Fisher's procedure to combine all of these P-values. Is there sufficient evidence to conclude that the population proportion exceeds 0.30?

9.11.2 Suppose that 10 sets of hypotheses of the form

$$H_0: \mu = \mu_0 \qquad H_1: \mu \neq \mu_0$$

have been tested and that the P-values for these tests are 0.12, 0.08. 0.93, 0.02, 0.01, 0.05, 0.88, 0.15, 0.13, and 0.06. Use Fisher's procedure to combine all of these P-values. What conclusions can you draw about these hypotheses?

9.11.3 SS The mean weight of a package of frozen fish must equal 22 oz. Five independent samples were selected, and the statistical hypotheses about the mean weight were tested. The P-values that resulted from these tests were 0.065, 0.0924, 0.073, 0.025, and 0.021. Is there sufficient evidence to conclude that the mean package weight is not equal to 22 oz?

9.11.4 The standard deviation of fill volume of a container of a pharmaceutical product must be less than 0.2 oz to ensure that the container is accurately filled. Six independent samples were selected, and the statistical hypotheses about the standard deviation were tested. The P-values that resulted were 0.15, 0.091, 0.075, 0.02, 0.04, and 0.06. Is there sufficient evidence to conclude that the standard deviation of fill volume is less than 0.2 oz?

Supplemental Exercises for Chapter 9

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

9.S5 WP Consider the following computer output.

One-Sample T:

```
Test of mu = 100 vs not = 100
Variable  N   Mean   StDev  SE Mean  95% CI  T  P
X         16  98.33  4.61      ?     (?, ?)  ?  ?
```

a. How many degrees of freedom are there on the t-statistic?
b. Fill in the missing information. You may use bounds on the P-value.
c. What are your conclusions if $\alpha = 0.05$?
d. What are your conclusions if the hypothesis is H_0: $\mu = 100$ versus H_0: $\mu > 100$?

9.S6 Consider the following computer output.

One-Sample Z:

```
Test of mu = 26 vs > 26
The assumed standard deviation = 1.5
Variable  N   Mean    StDev  SE Mean  Z     P
X         ?  26.541  2.032   0.401    ?     ?
```

a. Fill in the missing information.
b. Is this a one-sided or a two-sided test?
c. What are your conclusions if $\alpha = 0.05$?
d. Find a 95% two-sided CI on the mean.

9.S7 Consider the following computer output.

One-Sample T:

```
Test of mu = 85 vs < 85
Variable   N    Mean    StDev   SE Mean   T   P
X          25   84.331    ?      0.631    ?   ?
```

a. How many degrees of freedom are there on the t-statistic?
b. Fill in the missing information. You may use bounds on the P-value.
c. What are your conclusions if $\alpha = 0.05$?
d. Find a 95% upper-confidence bound on the mean.
e. What are your conclusions if the hypothesis is H_0: $\mu = 100$ versus H_0: $\mu > 100$?

9.S8 An article in *Fire Technology* ["An Experimental Examination of Dead Air Space for Smoke Alarms" (2009, Vol. 45, pp. 97–115)] studied the performance of smoke detectors installed not less than 100 mm from any adjoining wall if mounted on a flat ceiling, and not closer than 100 mm and not farther than 300 mm from the adjoining ceiling surface if mounted on walls. The purpose of this rule is to avoid installation of smoke alarms in the "dead air space," where it is assumed to be difficult for smoke to reach. The paper described a number of interesting experiments. Results on the time to signal (in seconds) for one such experiment with pine stick fuel in an open bedroom using photoelectric smoke alarms are as follows: 220, 225, 297, 315, 282, and 313.

a. Is there sufficient evidence to support a claim that the mean time to signal is less than 300 seconds?
b. Is there practical concern about the assumption of a normal distribution as a model for the time-to-signal data?
c. Find a 95% two-sided CI on the mean time to signal.

9.S9 SS Consider the situation in Exercise 9.S10. After collecting a sample, we are interested in testing H_0: $p = 0.10$ versus H_1: $p \neq 0.10$ with $\alpha = 0.05$. For each of the following situations, compute the p-value for this test:

a. $n = 50$, $\hat{p} = 0.095$
b. $n = 100$, $\hat{p} = 0.095$
c. $n = 500$, $\hat{p} = 0.095$
d. $n = 1000$, $\hat{p} = 0.095$
e. Comment on the effect of sample size on the observed P-value of the test.

9.S10 A manufacturer of semiconductor devices takes a random sample of size n of chips and tests them, classifying each chip as defective or nondefective. Let $X_i = 0$ if the chip is nondefective and $X_i = 1$ if the chip is defective. The sample fraction defective is

$$\hat{p} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

What are the sampling distribution, the sample mean, and sample variance estimates of $\hat{p}$ when

a. The sample size is $n = 50$?
b. The sample size is $n = 80$?
c. The sample size is $n = 100$?
d. Compare your answers to parts (a)–(c) and comment on the effect of sample size on the variance of the sampling distribution.

9.S11 Suppose that you are testing H_0: $p = 0.5$ versus H_0: $p \neq 0.5$. Suppose that p is the true value of the population proportion.

a. Using $\alpha = 0.05$, find the power of the test for $n = 100$, 150, and 300, assuming that $p = 0.6$. Comment on the effect of sample size on the power of the test.
b. Using $\alpha = 0.01$, find the power of the test for $n = 100$, 150, and 300, assuming that $p = 0.6$. Compare your answers to those from part (a) and comment on the effect of α on the power of the test for different sample sizes.
c. Using $\alpha = 0.05$, find the power of the test for $n = 100$, assuming $p = 0.08$. Compare your answer to part (a) and comment on the effect of the true value of p on the power of the test for the same sample size and α level.
d. Using $\alpha = 0.01$, what sample size is required if $p = 0.6$ and we want $\beta = 0.05$? What sample is required if $p = 0.8$ and we want $\beta = 0.05$? Compare the two sample sizes and

comment on the effect of the true value of p on a sample size required when β is held approximately constant.

9.S12 The cooling system in a nuclear submarine consists of an assembly of welded pipes through which a coolant is circulated. Specifications require that weld strength must meet or exceed 150 psi.

a. Suppose that the design engineers decide to test the hypothesis H_0: $\mu = 150$ versus H_1: $\mu > 150$. Explain why this choice of alternative hypothesis is better than H_1: $\mu < 150$.

b. A random sample of 20 welds results in $\bar{x} = 153.7$ psi and $s = 11.3$ psi. What conclusions can you draw about the hypothesis in part (a)? State any necessary assumptions about the underlying distribution of the data.

9.S13 A biotechnology company produces a therapeutic drug whose concentration has a standard deviation of 4 grams per liter. A new method of producing this drug has been proposed, although some additional cost is involved. Management will authorize a change in production technique only if the standard deviation of the concentration in the new process is less than 4 grams per liter. The researchers chose $n = 10$ and obtained the following data in grams per liter. Perform the necessary analysis to determine whether a change in production technique should be implemented.

16.628	16.630
16.622	16.631
16.627	16.624
16.623	16.622
16.618	16.626

9.S14 A manufacturer of precision measuring instruments claims that the standard deviation in the use of the instruments is at most 0.00002 millimeter. An analyst who is unaware of the claim uses the instrument eight times and obtains a sample standard deviation of 0.00001 millimeter.

a. Confirm using a test procedure and an α level of 0.01 that there is insufficient evidence to support the claim that the standard deviation of the instruments is at most 0.00002. State any necessary assumptions about the underlying distribution of the data.

b. Explain why the sample standard deviation, $s = 0.00001$, is less than 0.00002, yet the statistical test procedure results do not support the claim.

9.S15 SS Consider the water quality data in Exercise 6.2.10.

a. Do these data support the claim that the mean concentration of suspended solids does not exceed 50 parts per million? Use $\alpha = 0.05$.

b. What is the P-value for the test in part (a)?

c. Does the normal distribution seem to be a reasonable model for these data? Perform an appropriate goodness-of-fit test to answer this question.

9.S16 Consider the spot weld shear strength data in Exercise 6.2.9. Does the normal distribution seem to be a reasonable model for these data? Perform an appropriate goodness-of-fit test to answer this question.

9.S17 An article in *Food Testing and Analysis* ["Improving Reproducibility of Refractometry Measurements of Fruit Juices" (1999, Vol. 4(4), pp. 13–17)] measured the sugar concentration (Brix) in clear apple juice. All readings were taken at 20°C:

11.48	11.45	11.48	11.47	11.48
11.50	11.42	11.49	11.45	11.44
11.45	11.47	11.46	11.47	11.43
11.50	11.49	11.45	11.46	11.47

a. Test the hypothesis H_0: $\mu = 11.5$ versus H_1: $\mu \neq 11.5$ using $\alpha = 0.05$. Find the P-value.

b. Compute the power of the test if the true mean is 11.4.

c. What sample size would be required to detect a true mean sugar concentration of 11.45 if we wanted the power of the test to be at least 0.9?

d. Explain how the question in part (a) could be answered by constructing a two-sided confidence interval on the mean sugar concentration.

e. Is there evidence to support the assumption that the sugar concentration is normally distributed?

9.S18 Consider the computer output below.

Test and CI for One Proportion

Test of $p = 0.25$ vs $p < 0.25$

X	N	Sample p	Bound	Z-Value	P-Value
53	225	0.235556	0.282088	?	?

Using the normal approximation:

a. Fill in the missing information.

b. What are your conclusions if $\alpha = 0.05$?

c. The normal approximation to the binomial was used here. Was that appropriate?

d. Find a 95% upper-confidence bound on the true proportion.

e. What are the P-value and your conclusions if the alternative hypothesis is H_1: $p \neq 0.25$?

9.S19 An article in *Experimental Brain Research* ["Synapses in the Granule Cell Layer of the Rat Dentate Gyrus: Serial-Sectioning Study" (1996, Vol. 112(2), pp. 237–243)] showed the ratio between the numbers of symmetrical and total synapses on somata and azon initial segments of reconstructed granule cells in the dentate gyrus of a 12-week-old rat:

0.65	0.90	0.78	0.94	0.40	0.94
0.91	0.86	0.53	0.84	0.42	0.50
0.50	0.68	1.00	0.57	1.00	1.00
0.84	0.9	0.91	0.92	0.96	
0.96	0.56	0.67	0.96	0.52	
0.89	0.60	0.54			

a. Use the data to test H_0: $\sigma^2 = 0.02$ versus H_1: $\sigma^2 \neq 0.02$ using $\alpha = 0.05$.

b. Find the P-value for the test.

9.S20 An article in *Biological Trace Element Research* ["Interaction of Dietary Calcium, Manganese, and Manganese Source (Mn Oxide or Mn Methionine Complex) or Chick Performance and Manganese Utilization" (1991, Vol. 29(3), pp. 217–228)] showed the following results of tissue assay for liver manganese (ppm) in chicks fed high-Ca diets.

6.02	6.08	7.11	5.73	5.32	7.10
5.29	5.84	6.03	5.99	4.53	6.81

a. Test the hypothesis $H_0: \sigma^2 = 0.6$ versus $H_1: \sigma^2 \neq 0.6$ using $\alpha = 0.01$.

b. What is the *P*-value for this test?

c. Discuss how part (a) could be answered by constructing a 99% two-sided confidence interval for σ.

9.S21 Consider the television picture tube brightness experiment described in Exercise 8.2.7.

a. For the sample size $n = 10$, do the data support the claim that the standard deviation of current is less than 20 microamps?

b. Suppose that instead of $n = 10$, the sample size was 51. Repeat the analysis performed in part (a) using $n = 51$.

c. Compare your answers and comment on how sample size affects your conclusions drawn in parts (a) and (b).

Chapter 10 Exercises

Exercises for Section 10.1

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

10.1.1 Consider the hypothesis test $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$ with known variances $\sigma_1 = 10$ and $\sigma_2 = 5$. Suppose that sample sizes $n_1 = 10$ and $n_2 = 15$ and that $\bar{x}_1 = 4.7$ and $\bar{x}_2 = 7.8$. Use $\alpha = 0.05$.

a. Test the hypothesis and find the *P*-value.

b. Explain how the test could be conducted with a confidence interval.

c. What is the power of the test in part (a) for a true difference in means of 3?

d. Assume that sample sizes are equal. What sample size should be used to obtain $\beta = 0.05$ if the true difference in means is 3? Assume that $\alpha = 0.05$.

10.1.2 Two machines are used for filling plastic bottles with a net volume of 16.0 ounces. The fill volume can be assumed to be normal with standard deviation $\sigma_1 = 0.020$ and $\sigma_2 = 0.025$ ounces. A member of the quality engineering staff suspects that both machines fill to the same mean net volume, whether or not this volume is 16.0 ounces. A random sample of 10 bottles is taken from the output of each machine.

Machine 1		Machine 2	
16.03	16.01	16.02	16.03
16.04	15.96	15.97	16.04
16.05	15.98	15.96	16.02
16.05	16.02	16.01	16.01
16.02	15.99	15.99	16.00

a. Do you think the engineer is correct? Use $\alpha = 0.05$. What is the *P*-value for this test?

b. Calculate a 95% confidence interval on the difference in means. Provide a practical interpretation of this interval.

c. What is the power of the test in part (a) for a true difference in means of 0.04?

d. Assume that sample sizes are equal. What sample size should be used to ensure that $\beta = 0.05$ if the true difference in means is 0.04? Assume that $\alpha = 0.05$.

10.1.3 WP SS The burning rates of two different solid-fuel propellants used in air crew escape systems are being studied. It is known that both propellants have approximately the same standard deviation of burning rate; that is, $\sigma_1 = \sigma_2 = 3$ centimeters per second. Two random samples of $n_1 = 20$ and $n_2 = 20$ specimens are tested; the sample mean burning rates are $\bar{x}_1 = 18$ centimeters per second and $\bar{x}_2 = 24$ centimeters per second.

a. Test the hypothesis that both propellants have the same mean burning rate. Use $\alpha = 0.05$. What is the *P*-value?

b. Construct a 95% confidence interval on the difference in means $\mu_1 - \mu_2$. What is the practical meaning of this interval?

c. What is the β-error of the test in part (a) if the true difference in mean burning rate is 2.5 centimeters per second?

d. Assume that sample sizes are equal. What sample size is needed to obtain power of 0.9 at a true difference in means of 14 cm/s?

10.1.4 A polymer is manufactured in a batch chemical process. Viscosity measurements are normally made on each batch, and long experience with the process has indicated that the variability in the process is fairly stable with $\sigma = 20$. Fifteen batch viscosity measurements are given as follows:

724, 718, 776, 760, 745, 759, 795, 756, 742, 740, 761, 749, 739, 747, 742

A process change that involves switching the type of catalyst used in the process is made. Following the process change, eight batch viscosity measurements are taken:

735, 775, 729, 755, 783, 760, 738, 780

Assume that process variability is unaffected by the catalyst change. If the difference in mean batch viscosity is 10 or less, the manufacturer would like to detect it with a high probability.

a. Formulate and test an appropriate hypothesis using $\alpha = 0.10$. What are your conclusions? Find the P-value.

b. Find a 90% confidence interval on the difference in mean batch viscosity resulting from the process change.

c. Compare the results of parts (a) and (b) and discuss your findings.

10.1.5 An article in *Industrial Engineer* (September 2012) reported on a study of potential sources of injury to equine veterinarians conducted at a university veterinary hospital. Forces on the hand were measured for several common activities that veterinarians engage in when examining or treating horses. We consider the forces on the hands for two tasks, lifting and using ultrasound. Assume that both sample sizes are 6, the sample mean force for lifting was 6.0 pounds with standard deviation 1.5 pounds, and the sample mean force for using ultrasound was 6.2 pounds with standard deviation 0.3 pounds (data read from graphs in the article). Assume that the standard deviations are known. Is there evidence to conclude that the two activities result in significantly different forces on the hands?

Exercises for Section 10.2

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

10.2.1 Consider the following computer output.

Two-Sample T-Test and CI

Sample	N	Mean	StDev	SE Mean
1	12	10.94	1.26	0.36
2	16	12.15	1.99	0.50

```
Difference = mu (1) − mu (2)
Estimate for difference: −1.210
95% CI for difference: (−2.560, 0.140)
T-test of difference = 0 (vs not =) :
T-value = ? P-value = ? DF = ?
Both use Pooled StDev = ?
```

a. Fill in the missing values. Is this a one-sided or a two-sided test? Use lower and upper bounds for the P-value.

b. What are your conclusions if $\alpha = 0.05$? What if $\alpha = 0.01$?

c. This test was done assuming that the two population variances were equal. Does this seem reasonable?

d. Suppose that the hypothesis had been $H_0: \mu_1 = \mu_2$ versus $H_0: \mu_1 < \mu_2$. What would your conclusions be if $\alpha = 0.05$?

10.2.2 **WP** Consider the hypothesis test $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$. Suppose that sample sizes are $n_1 = 15$ and $n_2 = 15$, that $\bar{x}_1 = 4.7$ and $\bar{x}_2 = 7.8$, and that $s_1^2 = 4$ and $s_2^2 = 6.25$. Assume that $\sigma_1^2 = \sigma_2^2$ and that the data are drawn from normal distributions. Use $\alpha = 0.05$.

a. Test the hypothesis and find the P-value.

b. Explain how the test could be conducted with a confidence interval.

c. What is the power of the test in part (a) for a true difference in means of 3?

d. Assume that sample sizes are equal. What sample size should be used to obtain $\beta = 0.05$ if the true difference in means is -2? Assume that $\alpha = 0.05$.

10.2.3 **WP GO Tutorial** The diameter of steel rods manufactured on two different extrusion machines is being investigated. Two random samples of sizes $n_1 = 15$ and $n_2 = 17$ are selected, and the sample means and sample variances are $\bar{x}_1 = 8.73$, $s_1^2 = 0.35$, $\bar{x}_2 = 8.68$, and $s_2^2 = 0.40$, respectively. Assume that $\sigma_1^2 = \sigma_2^2$ and that the data are drawn from a normal distribution.

a. Is there evidence to support the claim that the two machines produce rods with different mean diameters? Use $\alpha = 0.05$ in arriving at this conclusion. Find the P-value.

b. Construct a 95% confidence interval for the difference in mean rod diameter. Interpret this interval.

10.2.4 **WP** Consider the hypothesis test $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$. Suppose that sample sizes $n_1 = 10$ and $n_2 = 10$, that $\bar{x}_1 = 7.8$ and $\bar{x}_2 = 5.6$, and that $s_1^2 = 4$ and $s_2^2 = 9$. Assume that $\sigma_1^2 = \sigma_2^2$ and that the data are drawn from normal distributions. Use $\alpha = 0.05$.

a. Test the hypothesis and find the P-value.

b. Explain how the test could be conducted with a confidence interval.

c. What is the power of the test in part (a) if μ_1 is 3 units greater than μ_2?

d. Assume that sample sizes are equal. What sample size should be used to obtain $\beta = 0.05$ if μ_1 is 3 units greater than μ_2? Assume that $\alpha = 0.05$.

10.2.5 **WP** Two catalysts may be used in a batch chemical process. Twelve batches were prepared using catalyst 1, resulting in an average yield of 86 and a sample standard deviation of 3. Fifteen batches were prepared using catalyst 2, and they resulted in an average yield of 89 with a standard deviation of 2. Assume that yield measurements are approximately normally distributed with the same standard deviation.

a. Is there evidence to support a claim that catalyst 2 produces a higher mean yield than catalyst 1? Use $\alpha = 0.01$.

b. Find a 99% confidence interval on the difference in mean yields that can be used to test the claim in part (a).

10.2.6 **WP** **VS** Two suppliers manufacture a plastic gear used in a laser printer. The impact strength of these gears measured in foot-pounds is an important characteristic. A random sample of 10 gears from supplier 1 results in $\bar{x}_1 = 290$ and $s_1 = 12$, and another random sample of 16 gears from the second supplier results in $\bar{x}_2 = 321$ and $s_2 = 22$.

a. Is there evidence to support the claim that supplier 2 provides gears with higher mean impact strength? Use

$\alpha = 0.05$, and assume that both populations are normally distributed but the variances are not equal. What is the *P*-value for this test?

b. Do the data support the claim that the mean impact strength of gears from supplier 2 is at least 25 foot-pounds higher than that of supplier 1? Make the same assumptions as in part (a).

c. Construct a confidence interval estimate for the difference in mean impact strength, and explain how this interval could be used to answer the question posed regarding supplier-to-supplier differences.

10.2.7 WP SS An article in *Electronic Components and Technology Conference* (2001, Vol. 52, pp. 1167–1171) compared single versus dual spindle saw processes for copper metallized wafers. A total of 15 devices of each type were measured for the width of the backside chipouts, $\bar{x}_{single} = 66.385$, $s_{single} = 7.895$ and $\bar{x}_{double} = 45.278$, $s_{double} = 8.612$.

a. Do the sample data support the claim that both processes have the same chip outputs? Use $\alpha = 0.05$ and assume that both populations are normally distributed and have the same variance. Find the *P*-value for the test.

b. Construct a 95% two-sided confidence interval on the mean difference in spindle saw process. Compare this interval to the results in part (a).

c. If the β-error of the test when the true difference in chip outputs is 15 should not exceed 0.1, what sample sizes must be used? Use $\alpha = 0.05$.

10.2.8 WP A photoconductor film is manufactured at a nominal thickness of 25 mils. The product engineer wishes to increase the mean speed of the film and believes that this can be achieved by reducing the thickness of the film to 20 mils. Eight samples of each film thickness are manufactured in a pilot production process, and the film speed (in microjoules per square inch) is measured. For the 25-mil film, the sample data result is $\bar{x}_1 = 1.15$ and $s_1 = 0.11$, and for the 20-mil film the data yield $\bar{x}_2 = 1.06$ and $s_2 = 0.09$. Note that an increase in film speed would lower the value of the observation in microjoules per square inch.

a. Do the data support the claim that reducing the film thickness increases the mean speed of the film? Use $\sigma = 0.10$, and assume that the two population variances are equal and the underlying population of film speed is normally distributed. What is the *P*-value for this test?

b. Find a 95% confidence interval on the difference in the two means that can be used to test the claim in part (a).

10.2.9 WP An article in *Radio Engineering and Electronic Physics* [1984, Vol. 29(3), pp. 63–66] investigated the behavior of a stochastic generator in the presence of external noise. The number of periods was measured in a sample of 100 trains for each of two different levels of noise voltage, 100 and 150 mV. For 100 mV, the mean number of periods in a train was 7.9 with $s = 2.6$. For 150 mV, the mean was 6.9 with $s = 2.4$.

a. It was originally suspected that raising noise voltage would reduce the mean number of periods. Do the data support this claim? Use $\alpha = 0.01$ and assume that each population is normally distributed and the two population variances are equal. What is the *P*-value for this test?

b. Calculate a confidence interval to answer the question in part (a).

10.2.10 VS An article in *IEEE International Symposium on Electromagnetic Compatibility* ["EM Effects of Different Mobile Handsets on Rats' Brain" (2002, Vol. 2, pp. 667–670)] quantified the absorption of electromagnetic energy and the resulting thermal effect from cellular phones. The experimental results were obtained from in vivo experiments conducted on rats. The arterial blood pressure values (mmHg) for the control group (8 rats) during the experiment are $\bar{x}_1 = 90$, $s_1 = 5$ and for the test group (9 rats) are $\bar{x}_2 = 115$, $s_2 = 10$.

a. Is there evidence to support the claim that the test group has higher mean blood pressure? Use $\alpha = 0.05$, and assume that both populations are normally distributed but the variances are not equal. What is the *P*-value for this test?

b. Calculate a confidence interval to answer the question in part (a).

c. Do the data support the claim that the mean blood pressure from the test group is at least 15 mmHg higher than the control group? Make the same assumptions as in part (a).

d. Explain how the question in part (c) could be answered with a confidence interval.

10.2.11 WP SS The overall distance traveled by a golf ball is tested by hitting the ball with Iron Byron, a mechanical golfer with a swing that is said to emulate the distance hit by the legendary champion Byron Nelson. Ten randomly selected balls of two different brands are tested and the overall distance measured. The data follow:

Brand 1: 275, 286, 287, 271, 283, 271, 279, 275, 263, 267

Brand 2: 258, 244, 260, 265, 273, 281, 271, 270, 263, 268

a. Is there evidence that overall distance is approximately normally distributed? Is an assumption of equal variances justified?

b. Test the hypothesis that both brands of ball have equal mean overall distance. Use $\alpha = 0.05$. What is the *P*-value?

c. Construct a 95% two-sided CI on the mean difference in overall distance for the two brands of golf balls.

d. What is the power of the statistical test in part (b) to detect a true difference in mean overall distance of 5 yards?

e. What sample size would be required to detect a true difference in mean overall distance of 3 yards with power of approximately 0.75?

10.2.12 The "spring-like effect" in a golf club could be determined by measuring the coefficient of restitution (the ratio of the outbound velocity to the inbound velocity of a golf ball fired at the clubhead). Twelve randomly selected drivers produced by two clubmakers are tested and the coefficient of restitution measured. The data follow:

Club 1: 0.8406, 0.8104, 0.8234, 0.8198, 0.8235, 0.8562, 0.8123, 0.7976, 0.8184, 0.8265, 0.7773, 0.7871

Club 2: 0.8305, 0.7905, 0.8352, 0.8380, 0.8145, 0.8465, 0.8244, 0.8014, 0.8309, 0.8405, 0.8256, 0.8476

a. Is there evidence that coefficient of restitution is approximately normally distributed? Is an assumption of equal variances justified?

b. Test the hypothesis that both brands of clubs have equal mean coefficient of restitution. Use $\alpha = 0.05$. What is the P-value of the test?
c. Construct a 95% two-sided CI on the mean difference in coefficient of restitution for the two brands of golf clubs.
d. What is the power of the statistical test in part (b) to detect a true difference in mean coefficient of restitution of 0.2?
e. What sample size would be required to detect a true difference in mean coefficient of restitution of 0.1 with power of approximately 0.8?

10.2.13 A paper in *Quality Engineering* [2013, Vol. 25(1)] presented data on cycles to failure of solder joints at different temperatures for different types of printed circuit boards (PCB). Failure data for two temperatures (20 and 60°C) for a copper-nickel-gold PCB follow.

20°C	218, 265, 279, 282, 336, 469, 496, 507, 685, 685
60°C	185, 242, 254, 280, 305, 353, 381, 504, 556, 697

a. Test the null hypothesis at $\alpha = 0.05$ that the cycles to failure are the same at both temperatures. Is the alternative one or two sided?
b. Find a 95% confidence interval for the difference in the mean cycles to failure for the two temperatures.
c. Is the value zero contained in the 95% confidence interval? Explain the connection with the conclusion you reached in part (a).
d. Do normal probability plots of part cycles to failure indicate any violations of the assumptions for the tests and confidence interval that you performed?

10.2.14 An article in *Quality Engineering* [2012, Vol. 24(1)] described an experiment on a grinding wheel. The following are some of the grinding force data (in N) from this experiment at two different vibration levels.

Low 242, 249, 235, 250, 254, 244, 258, 311, 237, 261, 314, 252
High 302, 421, 419, 399, 317, 311, 350, 363, 392, 367, 301, 302

a. Is there evidence to support the claim that the mean grinding force increases with the vibration level?
b. Find a 95% confidence interval for the difference in the mean grinding force for the two vibration levels.
c. Is the value zero contained in the 95% confidence interval? Explain the connection with the conclusion you reached in part (a).
d. Do normal probability plots of grinding force indicate any violations of the assumptions for the tests and confidence interval that you performed?

Exercises for Section 10.3

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

10.3.1 WP An electrical engineer must design a circuit to deliver the maximum amount of current to a display tube to achieve sufficient image brightness. Within her allowable design constraints, she has developed two candidate circuits and tests prototypes of each. The resulting data (in microamperes) are as follows:

Circuit 1:	251, 255, 258, 257, 250, 251, 254, 250, 248
Circuit 2:	250, 253, 249, 256, 259, 252, 260, 251

a. Use the Wilcoxon rank-sum test to test $H_0: \mu_1 = \mu_2$ against the alternative $H_1: \mu_1 > \mu_2$. Use $\alpha = 0.025$.
b. Use the normal approximation for the Wilcoxon rank-sum test. Assume that $\alpha = 0.05$. Find the approximate P-value for this test statistic.

10.3.2 WP One of the authors travels regularly to Seattle, Washington. He uses either Delta or Alaska Airlines. Flight delays are sometimes unavoidable, but he would be willing to give most of his business to the airline with the best on-time arrival record. The number of minutes that his flight arrived late for the last six trips on each airline follows. Is there evidence that either airline has superior on-time arrival performance? Use $\alpha = 0.01$ and the Wilcoxon rank-sum test.

Delta:	13, 10, 1, 4, 0, 9 (minutes late)
Alaska:	15, 8, 3, −1, −2, 4 (minutes late)

10.3.3 WP SS VS The manufacturer of a hot tub is interested in testing two different heating elements for its product. The element that produces the maximum heat gain after 15 minutes would be preferable. The manufacturer obtains 10 samples of each heating unit and tests each one. The heat gain after 15 minutes (in °F) follows.

Unit 1:	25, 27, 29, 31, 30, 26, 24, 32, 33, 38
Unit 2:	31, 33, 32, 35, 34, 29, 38, 35, 37, 30

a. Is there any reason to suspect that one unit is superior to the other? Use $\alpha = 0.05$ and the Wilcoxon rank-sum test.
b. Use the normal approximation for the Wilcoxon rank-sum test. Assume that $\alpha = 0.05$. What is the approximate P-value for this test statistic?

10.3.4 Consider the distance traveled by a golf ball in Exercise 10.2.11.

a. Use the Wilcoxon rank-sum test to investigate if the means differ. Use $\alpha = 0.05$.

b. Use the normal approximation for the Wilcoxon rank-sum test with $\alpha = 0.05$. Find the approximate P-value for this test.

Exercises for Section 10.4

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

10.4.1 WP The manager of a fleet of automobiles is testing two brands of radial tires and assigns one tire of each brand at random to the two rear wheels of eight cars and runs the cars until the tires wear out. The data (in kilometers) follow. Find a 99% confidence interval on the difference in mean life. Which brand would you prefer based on this calculation?

Car	Brand 1	Brand 2
1	36,925	34,318
2	45,300	42,280
3	36,240	35,500
4	32,100	31,950
5	37,210	38,015
6	48,360	47,800
7	38,200	37,810
8	33,500	33,215

10.4.2 WP VS A computer scientist is investigating the usefulness of two different design languages in improving programming tasks. Twelve expert programmers who are familiar with both languages are asked to code a standard function in both languages and the time (in minutes) is recorded. The data follow:

	Time	
Programmer	Design Language 1	Design Language 2
1	17	18
2	16	14
3	21	19
4	14	11
5	18	23
6	24	21
7	16	10
8	14	13
9	21	19
10	23	24
11	13	15
12	18	20

a. Is the assumption that the difference in coding time is normally distributed reasonable?

b. Find a 95% confidence interval on the difference in mean coding times. Is there any indication that one design language is preferable?

10.4.3 WP An article in the *Journal of Aircraft* ["Equivalent Plate Analysis of Aircraft Wing Box Structures with General Planform Geometry" (1986, Vol. 23, pp. 859–864)] described a new equivalent plate analysis method formulation that is capable of modeling aircraft structures such as cranked wing boxes and that produces results similar to the more computationally intensive finite element analysis method. Natural vibration frequencies for the cranked wing box structure are calculated using both methods, and results for the first seven natural frequencies follow:

Freq.	Finite Element, Cycle/s	Equivalent Plate, Cycle/s
1	14.58	14.76
2	48.52	49.10
3	97.22	99.99
4	113.99	117.53
5	174.73	181.22
6	212.72	220.14
7	277.38	294.80

a. Do the data suggest that the two methods provide the same mean value for natural vibration frequency? Use $\alpha = 0.05$. Find the P-value.

b. Find a 95% confidence interval on the mean difference between the two methods.

10.4.4 WP Two different analytical tests can be used to determine the impurity level in steel alloys. Eight specimens are tested using both procedures, and the results are shown in the following tabulation.

Specimen	Test 1	Test 2
1	1.2	1.4
2	1.3	1.7
3	1.5	1.5
4	1.4	1.3
5	1.7	2.0
6	1.8	2.1
7	1.4	1.7
8	1.3	1.6

a. Is there sufficient evidence to conclude that tests differ in the mean impurity level, using $\alpha = 0.01$?

b. Is there evidence to support the claim that test 1 generates a mean difference 0.1 units lower than test 2? Use $\alpha = 0.05$.

c. If the mean from test 1 is 0.1 less than the mean from test 2, it is important to detect this with probability at least 0.90. Was the use of eight alloys an adequate sample size? If not, how many alloys should have been used?

10.4.5 SS Neuroscientists conducted research in a Canadian prison to see whether solitary confinement affects brain wave activity ["Changes in EEG Alpha Frequency and Evoked Response Latency During Solitary Confinement," *Journal of Abnormal Psychology* (1972, Vol. 7, pp. 54–59)]. They randomly assigned 20 inmates to two groups, assigning half to solitary confinement and the other half to regular confinement. The data follow:

Nonconfined	Confined
10.7	9.6
10.7	10.4
10.4	9.7
10.9	10.3
10.5	9.2
10.3	9.3
9.6	9.9
11.1	9.5
11.2	9.0
10.4	10.9

a. Is a paired *t*-test appropriate for testing whether the mean alpha wave frequencies are the same in the two groups? Explain.

b. Perform an appropriate test.

10.4.6 In *Biometrics* (1990, Vol. 46, pp. 673–87), the authors analyzed the circumference of five orange trees (labeled as A–E) measured on seven occasions (x_i).

Tree	x_1	x_2	x_3	x_4	x_5	x_6	x_7
A	30	58	87	115	120	142	145
B	33	69	111	156	172	203	203
C	30	51	75	108	115	139	140
D	32	62	112	167	179	209	214
E	30	49	81	125	142	174	177

a. Compare the mean increase in circumference in periods 1 to 2 to the mean increase in periods 2 to 3. The increase is the difference in circumference in the two periods. Are these means significantly different at $\alpha = 0.10$?

b. Is there evidence that the mean increase in period 1 to period 2 is greater than the mean increase in period 6 to period 7 at $\alpha = 0.05$?

c. Are the assumptions of the test in part (a) violated because the same data (period 2 circumference) are used to calculate both mean increases?

10.4.7 In a series of tests to study the efficacy of ginkgo biloba on memory, Solomon et al. first looked at differences in memory tests of people 6 weeks before and after joining the study ["Ginkgo for Memory Enhancement: A Randomized Controlled Trial," *Journal of the American Medical Association* (2002, Vol. 288, pp. 835–840)]. For 99 patients receiving no medication, the average increase in category fluency (number of words generated in one minute) was 1.07 words with a standard deviation of 3.195 words. Researchers wanted to know whether the mean number of words recalled was positive.

a. Is this a one- or two-sided test?

b. Perform a hypothesis test to determine whether the mean increase is zero.

c. Why can this be viewed as a paired *t*-test?

d. What does the conclusion say about the importance of including placebos in such tests?

Exercises for Section 10.5

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

10.5.1 For an *F* distribution, find the following:

a. $f_{0.25,5,10}$ **b.** $f_{0.10,24,9}$
c. $f_{0.05,8,15}$ **d.** $f_{0.75,5,10}$
e. $f_{0.90,24,9}$ **f.** $f_{0.95,8,15}$

10.5.2 WP Two chemical companies can supply a raw material. The concentration of a particular element in this material is important. The mean concentration for both suppliers is the same, but you suspect that the variability in concentration may differ for the two companies. The standard deviation of concentration in a random sample of $n_1 = 10$ batches produced by company 1 is $s_1 = 4.7$ grams per liter, and for company 2, a random sample of $n_2 = 16$ batches yields $s_2 = 5.8$ grams per liter. Is there sufficient evidence to conclude that the two population variances differ? Use $\alpha = 0.05$.

10.5.3 Consider the hypothesis test $H_0: \sigma_1^2 = \sigma_2^2$ against $H_1: \sigma_1^2 < \sigma_2^2$, respectively. Suppose that the sample sizes are $n_1 = 5$ and $n_2 = 10$, and that $s_1^2 = 23.2$ and $s_2^2 = 28.8$. Use $\alpha = 0.05$. Test the hypothesis and explain how the test could be conducted with a confidence interval on σ_1/σ_2.

10.5.4 WP Consider the gear impact strength data in Exercise 10.2.6. Is there sufficient evidence to conclude that the variance of impact strength is different for the two suppliers? Use $\alpha = 0.05$.

10.5.5 WP SS Reconsider the overall distance data for golf balls in Exercise 10.2.11. Is there evidence to support the claim that the standard deviation of overall distance is the same for both brands of balls (use $\alpha = 0.05$)? Explain how this question can be answered with a 95% confidence interval on σ_1/σ_2.

10.5.6 Reconsider the coefficient of restitution data in Exercise 10.2.12. Do the data suggest that the standard deviation is the same for both brands of drivers (use $\alpha = 0.05$)? Explain how to answer this question with a confidence interval on σ_1/σ_2.

Exercises for Section 10.6

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

10.6.1 Consider the following computer output.

Test and CI for Two Proportions

Sample	X	N	Sample p
1	54	250	0.216000
2	60	290	0.206897

```
Difference = p(1) –p(2)
Estimate for difference: 0.00910345
95% CI for difference: (-0.0600031,
0.0782100)
Test for difference
= 0(vs not=0):Z = ? P-Value = ?
```

a. Is this a one-sided or a two-sided test?
b. Fill in the missing values.
c. Can the null hypothesis be rejected?
d. Construct an approximate 90% CI for the difference in the two proportions.

10.6.2 WP An article in *Knee Surgery, Sports Traumatology, Arthroscopy* (2005, Vol. 13(4), pp. 273–279) considered arthroscopic meniscal repair with an absorbable screw. Results showed that for tears greater than 25 millimeters, 14 of 18 (78%) repairs were successful, but for shorter tears, 22 of 30 (73%) repairs were successful.

a. Is there evidence that the success rate is greater for longer tears? Use $\alpha = 0.05$. What is the *P*-value?
b. Calculate a one-sided 95% confidence bound on the difference in proportions that can be used to answer the question in part (a).

10.6.3 SS VS Two different types of injection-molding machines are used to form plastic parts. A part is considered defective if it has excessive shrinkage or is discolored. Two random samples, each of size 300, are selected, and 15 defective parts are found in the sample from machine 1, and 8 defective parts are found in the sample from machine 2.

a. Is it reasonable to conclude that both machines produce the same fraction of defective parts, using $\alpha = 0.05$? Find the *P*-value for this test.
b. Construct a 95% confidence interval on the difference in the two fractions defective.
c. Suppose that $p_1 = 0.05$ and $p_2 = 0.01$. With the sample sizes given here, what is the power of the test for this two-sided alternate?
d. Suppose that $p_1 = 0.05$ and $p_2 = 0.01$. Determine the sample size needed to detect this difference with a probability of at least 0.9.
e. Suppose that $p_1 = 0.05$ and $p_2 = 0.02$. With the sample sizes given here, what is the power of the test for this two-sided alternate?
f. Suppose that $p_1 = 0.05$ and $p_2 = 0.02$. Determine the sample size needed to detect this difference with a probability of at least 0.9.

10.6.4 A random sample of 500 adult residents of Maricopa County indicated that 385 were in favor of increasing the highway speed limit to 75 mph, and another sample of 400 adult residents of Pima County indicated that 267 were in favor of the increased speed limit.

a. Do these data indicate that there is a difference in the support for increasing the speed limit for the residents of the two counties? Use $\alpha = 0.05$. What is the *P*-value for this test?
b. Construct a 95% confidence interval on the difference in the two proportions. Provide a practical interpretation of this interval.

10.6.5 The *New England Journal of Medicine* reported an experiment to judge the efficacy of surgery on men diagnosed with prostate cancer. The randomly assigned half of 695 (347) men in the study had surgery, and 18 of them eventually died of prostate cancer compared with 31 of the 348 who did not have surgery. Is there any evidence to suggest that the surgery lowered the proportion of those who died of prostate cancer?

10.6.6 Consider the highway speed limit data introduced in Exercise 10.6.4. Find a 99% CI on the difference in the two proportions using the alternate CI procedure described in this section. Compare the lengths of the CI from Exercise 10.6.4 with those in this one. Discuss the possible causes of any differences that you observe.

Supplemental Exercises for Chapter 10

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

10.S7 Consider the following computer output.

Two-Sample T-Test and CI

Sample	N	Mean	StDev	SE Mean
1	20	11.87	2.23	?
2	20	12.73	3.19	

```
0.71 Difference = mu(1) - mu(2)
Estimate for difference: −0.860
95% CI for difference: (?, ?)
T-Test of difference = 0(vs not =):
T-Value = ? P-Value = ? DF = ?
Both use Pooled StDev = ?
```

a. Fill in the missing values. You may use bounds for the *P*-value.
b. Is this a two-sided test or a one-sided test?
c. What are your conclusions if $\alpha = 0.05$? What if $\alpha = 0.10$?

10.S8 A procurement specialist has purchased 25 resistors from vendor 1 and 35 resistors from vendor 2. Each resistor's resistance is measured with the following results (ohm):

Vendor 1

96.8	100.0	100.3	98.5	98.3	98.2
99.6	99.4	99.9	101.1	103.7	97.7
99.7	101.1	97.7	98.6	101.9	101.0
99.4	99.8	99.1	99.6	101.2	98.2
98.6					

Vendor 2

106.8	106.8	104.7	104.7	108.0	102.2
103.2	103.7	106.8	105.1	104.0	106.2
102.6	100.3	104.0	107.0	104.3	105.8
104.0	106.3	102.2	102.8	104.2	103.4
104.6	103.5	106.3	109.2	107.2	105.4
106.4	106.8	104.1	107.1	107.7	

a. What distributional assumption is needed to test the claim that the variance of resistance of the product from vendor 1 is not significantly different from the variance of resistance of the product from vendor 2? Perform a graphical procedure to check this assumption.
b. Perform an appropriate statistical hypothesis-testing procedure to determine whether the procurement specialist can claim that the variance of resistance of the product from vendor 1 is significantly different from the variance of resistance of the product from vendor 2.

10.S9 SS A liquid dietary product implies in its advertising that using the product for one month results in an average weight loss of at least 3 pounds. Eight subjects use the product for one month, and the resulting weight loss data follow. Use hypothesis-testing procedures to answer the following questions.

Subject	Initial Weight (lb)	Final Weight (lb)
1	165	161
2	201	195
3	195	192
4	198	193
5	155	150
6	143	141
7	150	146
8	187	183

a. Do the data support the claim of the dietary product's producer with the probability of a type I error set to 0.05?
b. Do the data support the claim of the dietary product's producer with the probability of a type I error set to 0.01?
c. In an effort to improve sales, the producer is considering changing its claim from "at least 3 pounds" to "at least 5 pounds." Repeat parts (a) and (b) to test this new claim.

10.S10 The Salk polio vaccine experiment in 1954 focused on the effectiveness of the vaccine in combating paralytic polio. Because it was believed that without a control group of children, there would be no sound basis for evaluating the efficacy of the Salk vaccine, the vaccine was administered to one group, and a placebo (visually identical to the vaccine but known to have no effect) was administered to a second group. For ethical reasons and because it was suspected that knowledge of vaccine administration would affect subsequent diagnoses, the experiment was conducted in a double-blind fashion. That is, neither the subjects nor the administrators knew who received the vaccine and who received the placebo. The actual data for this experiment are as follows:

Placebo group: $n = 201{,}299$: 110 cases of polio observed
Vaccine group: $n = 200{,}745$: 33 cases of polio observed

a. Use a hypothesis-testing procedure to determine whether the proportion of children in the two groups who contracted paralytic polio is statistically different. Use a probability of a type I error equal to 0.05.
b. Repeat part (a) using a probability of a type I error equal to 0.01.
c. Compare your conclusions from parts (a) and (b) and explain why they are the same or different.

10.S11 WP A random sample of 1500 residential telephones in Phoenix in 1990 indicated that 387 of the numbers were unlisted. A random sample of 1200 telephones in the same year in Scottsdale indicated that 310 were unlisted.

a. Find a 95% confidence interval on the difference in the two proportions and use this confidence interval to determine whether there is a statistically significant difference in proportions of unlisted numbers between the two cities.

b. Find a 90% confidence interval on the difference in the two proportions and use this confidence interval to determine if there is a statistically significant difference in proportions of unlisted numbers for the two cities.

c. Suppose that all the numbers in the problem description were doubled. That is, 774 residents of 3000 sampled in Phoenix and 620 residents of 2400 in Scottsdale had unlisted phone numbers. Repeat parts (a) and (b) and comment on the effect of increasing the sample size without changing the proportions on your results.

10.S12 The manufacturer of a new pain relief tablet would like to demonstrate that its product works twice as fast as the competitor's product. Specifically, the manufacturer would like to test

$$H_0: \mu_1 = 2\mu_2$$
$$H_1: \mu_1 > 2\mu_2$$

where μ_1 is the mean absorption time of the competitive product and μ_2 is the mean absorption time of the new product. Assuming that the variances σ_1^2 and σ_2^2 are known, develop a procedure for testing this hypothesis.

10.S13 WP A fuel-economy study was conducted for two German automobiles, Mercedes and Volkswagen. One vehicle of each brand was selected, and the mileage performance was observed for 10 tanks of fuel in each car. The data are as follows (in miles per gallon):

Mercedes		Volkswagen	
24.7	24.9	41.7	42.8
24.8	24.6	42.3	42.4
24.9	23.9	41.6	39.9
24.7	24.9	39.5	40.8
24.5	24.8	41.9	29.6

a. Construct a normal probability plot of each of the data sets. Based on these plots, is it reasonable to assume that they are each drawn from a normal population?

b. Suppose that it was determined that the lowest observation of the Mercedes data was erroneously recorded and should be 24.6. Furthermore, the lowest observation of the Volkswagen data was also mistaken and should be 39.6. Again construct normal probability plots of each of the data sets with the corrected values. Based on these new plots, is it reasonable to assume that each is drawn from a normal population?

c. Compare your answers from parts (a) and (b) and comment on the effect of these mistaken observations on the normality assumption.

d. Using the corrected data from part (b) and a 95% confidence interval, is there evidence to support the claim that the variability in mileage performance is greater for a Volkswagen than for a Mercedes?

e. Rework part (d) of this problem using an appropriate hypothesis-testing procedure. Did you get the same answer as you did originally? Why?

10.S14 WP Suppose that you are testing $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \neq \mu_2$, and you plan to use equal sample sizes from the two populations. Both populations are assumed to be normal with unknown but equal variances. If you use $\alpha = 0.05$ and if the true mean $\mu_1 = \mu_2 + \sigma$, what sample size must be used for the power of this test to be at least 0.90?

10.S15 WP A Rockwell hardness-testing machine presses a tip into a test coupon and uses the depth of the resulting depression to indicate hardness. Two different tips are being compared to determine whether each provides the same Rockwell C-scale hardness readings. Nine coupons are tested with both tips being tested on each coupon. The data are shown in the following table.

a. State any assumptions necessary to test the claim that each tip produces the same Rockwell C-scale hardness readings. Check those assumptions for which you have the information.

Coupon	Tip 1	Tip 2	Coupon	Tip 1	Tip 2
1	47	46	6	41	41
2	42	40	7	45	46
3	43	45	8	45	46
4	40	41	9	49	48
5	42	43			

b. Apply an appropriate statistical method to determine whether the data support the claim that the difference in Rockwell C-scale hardness readings of the two tips differ significantly from zero.

c. Suppose that if the two tips differ in mean hardness readings by as much as 1.0, you want the power of the test to be at least 0.9. For an $\alpha = 0.01$, how many coupons should have been used in the test?

10.S16 WP An article in the *Journal of the Environmental Engineering Division* ["Distribution of Toxic Substances in Rivers" (1982, Vol. 108, pp. 639–649)] investigated the concentration of several hydrophobic organic substances in the Wolf River in Tennessee. Measurements on hexachlorobenzene (HCB) in nanograms per liter were taken at different depths downstream of an abandoned dump site. Data for two depths follow:

Surface: 3.74, 4.61, 4.00, 4.67, 4.87, 5.12, 4.52, 5.29, 5.74, 5.48
Bottom: 5.44, 6.88, 5.37, 5.44, 5.03, 6.48, 3.89, 5.85, 6.85, 7.16

a. What assumptions are required to test the claim that mean HCB concentration is the same at both depths? Check those assumptions for which you have the information.

b. Apply an appropriate procedure to determine whether the data support the claim in part (a).
c. Suppose that the true difference in mean concentrations is 2.0 nanograms per liter. For $\alpha = 0.05$, what is the power of a statistical test for H_0: $\mu_1 = \mu_2$ versus H_1: $\mu_1 \neq \mu_2$?
d. What sample size would be required to detect a difference of 1.0 nanograms per liter at $\alpha = 0.05$ if the power must be at least 0.9?

10.S17 Consider the unlisted telephone number data in Exercise 10.S11. Find 95% CIs on the difference in the proportions of unlisted telephone numbers for Phoenix and Scottsdale residents using both procedures described in this chapter. Compare the lengths of these two intervals and comment on any difference you may observe.

Chapter 11 Exercises

Exercises for Section 11.2

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

11.2.1 Diabetes and obesity are serious health concerns in the United States and much of the developed world. Measuring the amount of body fat a person carries is one way to monitor weight control progress, but measuring it accurately involves either expensive X-ray equipment or a pool in which to dunk the subject. Instead body mass index (BMI) is often used as a proxy for body fat because it is easy to measure: BMI = mass(kg)/(height(m))^2 = 703 mass(lb)/(height(in))^2. In a study of 250 men at Brigham Young University, both BMI and body fat were measured. Researchers found the following summary statistics:

$$\sum_{i=1}^{n} x_i = 6322.28 \quad \sum_{i=1}^{n} x_i^2 = 162674.18$$

$$\sum_{i=1}^{n} y_i = 4757.90 \quad \sum_{i=1}^{n} y_i^2 = 107679.27$$

$$\sum_{i=1}^{n} x_i y_i = 125471.10$$

a. **WP** Calculate the least squares estimates of the slope and intercept. Graph the regression line.
b. **WP** Use the equation of the fitted line to predict what body fat would be observed, on average, for a man with a BMI of 30.
c. Suppose that the observed body fat of a man with a BMI of 25 is 25%. Find the residual for that observation.
d. Was the prediction for the BMI of 25 in part (c) an overestimate or underestimate? Explain briefly.

11.2.2 **WP** An article in *Technometrics* by S. C. Narula and J. F. Wellington ["Prediction, Linear Regression, and a Minimum Sum of Relative Errors" (1977, Vol. 19(2), pp. 185–190)] presents data on the selling price and annual taxes for 24 houses. The data are in the table that follows.

a. Assuming that a simple linear regression model is appropriate, obtain the least squares fit relating selling price to taxes paid. What is the estimate of σ^2?
b. Find the mean selling price given that the taxes paid are $x = 7.50$.
c. Calculate the fitted value of y corresponding to $x = 5.8980$. Find the corresponding residual.
d. Calculate the fitted $\hat{y}_i$ for each value of x_i used to fit the model. Then construct a graph of $\hat{y}_i$ versus the corresponding observed value y_i and comment on what this plot would look like if the relationship between y and x was a deterministic (no random error) straight line. Does the plot actually obtained indicate that taxes paid is an effective regressor variable in predicting selling price?

Sale Price/ 1000	Taxes (local, school, county)/1000	Sale Price/ 1000	Taxes (local, school, county)/1000
25.9	4.9176	30.0	5.0500
29.5	5.0208	36.9	8.2464
27.9	4.5429	41.9	6.6969
25.9	4.5573	40.5	7.7841
29.9	5.0597	43.9	9.0384
29.9	3.8910	37.5	5.9894
30.9	5.8980	37.9	7.5422
28.9	5.6039	44.5	8.7951
35.9	5.8282	37.9	6.0831
31.5	5.3003	38.9	8.3607
31.0	6.2712	36.9	8.1400
30.9	5.9592	45.8	9.1416

11.2.3 WP An article in *Concrete Research* ["Near Surface Characteristics of Concrete: Intrinsic Permeability" (1989, Vol. 41, pp. 87–97)] presented data on compressive strength x and intrinsic permeability y of various concrete mixes and cures. Summary quantities are $n = 14$, $\sum y_i = 572$, $\sum y_i^2 = 23{,}530$, $\sum x_i = 43$, $\sum x_i^2 = 157.42$, and $\sum x_i y_i = 1697.80$. Assume that the two variables are related according to the simple linear regression model.

a. Calculate the least squares estimates of the slope and intercept. Estimate σ^2. Graph the regression line.

b. Use the equation of the fitted line to predict what permeability would be observed when the compressive strength is $x = 4.3$.

c. Give a point estimate of the mean permeability when the compressive strength is $x = 3.7$.

d. Suppose that the observed value of permeability at $x = 3.7$ is $y = 46.1$. Calculate the value of the corresponding residual.

11.2.4 An article in the *Journal of Sound and Vibration* ["Measurement of Noise-Evoked Blood Pressure by Means of Averaging Method: Relation between Blood Pressure Rise and *SPL*" (1991, Vol. 151(3), pp. 383–394)] described a study investigating the relationship between noise exposure and hypertension. The following data are representative of those reported in the article.

y	1	0	1	2	5	1	4	6	2	3
x	60	63	65	70	70	70	80	90	80	80
y	5	4	6	8	4	5	7	9	7	6
x	85	89	90	90	90	90	94	100	100	100

a. Draw a scatter diagram of y (blood pressure rise in millimeters of mercury) versus x (sound pressure level in decibels). Does a simple linear regression model seem reasonable in this situation?

b. Fit the simple linear regression model using least squares. Find an estimate of σ^2.

c. Find the predicted mean rise in blood pressure level associated with a sound pressure level of 85 decibels.

11.2.5 SS An article in the *Tappi Journal* (March 1986) presented data on green liquor Na_2S concentration (in grams per liter) and paper machine production (in tons per day). The data (read from a graph) follow:

y	40	42	49	46	44	48	
x	825	830	890	895	890	910	
y	46	43	53	52	54	57	58
x	915	960	990	1010	1012	1030	1050

a. Fit a simple linear regression model with y = green liquor Na_2S concentration and x = production. Find an estimate of σ^2. Draw a scatter diagram of the data and the resulting least squares fitted model.

b. Find the fitted value of y corresponding to $x = 910$ and the associated residual.

c. Find the mean green liquor Na_2S concentration when the production rate is 950 tons per day.

11.2.6 WP VS An article in the *Journal of Environmental Engineering* (1989, Vol. 115(3), pp. 608–619) reported the results of a study on the occurrence of sodium and chloride in surface streams in central Rhode Island. The following data are chloride concentration y (in milligrams per liter) and roadway area in the watershed x (in percentage).

y	4.4	6.6	9.7	10.6	10.8	10.9
x	0.19	0.15	0.57	0.70	0.67	0.63
y	11.8	12.1	14.3	14.7	15.0	17.3
x	0.47	0.70	0.60	0.78	0.81	0.78
y	19.2	23.1	27.4	27.7	31.8	39.5
x	0.69	1.30	1.05	1.06	1.74	1.62

a. Draw a scatter diagram of the data. Does a simple linear regression model seem appropriate here?

b. Fit the simple linear regression model using the method of least squares. Find an estimate of σ^2.

c. Estimate the mean chloride concentration for a watershed that has 1% roadway area.

d. Find the fitted value corresponding to $x = 0.47$ and the associated residual.

11.2.7 WP GO Tutorial SS An article in the *Journal of the American Ceramic Society* ["Rapid Hot-Pressing of Ultrafine PSZ Powders" (1991, Vol. 74, pp. 1547–1553)] considered the microstructure of the ultrafine powder of partially stabilized zirconia as a function of temperature. The data follow:

x = Temperature (°C):	1100	1200	1300	1100	1500
	1200	1300			
y = Porosity (%):	30.8	19.2	6.0	13.5	11.4
	7.7	3.6			

a. Fit the simple linear regression model using the method of least squares. Find an estimate of σ^2.

b. Estimate the mean porosity for a temperature of 1400 °C.

c. Find the fitted value corresponding to $y = 11.4$ and the associated residual.

d. Draw a scatter diagram of the data. Does a simple linear regression model seem appropriate here? Explain.

11.2.8 An article in *Wood Science and Technology* ["Creep in Chipboard, Part 3: Initial Assessment of the Influence of Moisture Content and Level of Stressing on Rate of Creep and Time to Failure" (1981, Vol. 15, pp. 125–144)] reported a study of the deflection (mm) of particleboard from stress levels of relative humidity. Assume that the two variables are related according to the simple linear regression model. The data follow:

x = Stress level (%):	54	54	61	61	68
y = Deflection (mm):	16.473	18.693	14.305	15.121	13.505
x = Stress level (%):	68	75	75	75	
y = Deflection (mm):	11.640	11.168	12.534	11.224	

a. Calculate the least square estimates of the slope and intercept. What is the estimate of σ^2? Graph the regression model and the data.

b. Find the estimate of the mean deflection if the stress level can be limited to 65%.

c. Estimate the change in the mean deflection associated with a 5% increment in stress level.

d. To decrease the mean deflection by one millimeter, how much increase in stress level must be generated?

e. Given that the stress level is 68%, find the fitted value of deflection and the corresponding residual.

11.2.9 WP An article in the *Journal of the Environmental Engineering Division* ["Least Squares Estimates of BOD Parameters" (1980, Vol. 106, pp. 1197–1202)] took a sample from the Holston River below Kingport, Tennessee, during August 1977. The biochemical oxygen demand (BOD) test is conducted over a period of time in days. The resulting data follow:

Time (days):	1	2	4	6	8	10	12	14	16
	18	20							
BOD (mg/liter):	0.6	0.7	1.5	1.9	2.1	2.6	2.9	3.7	3.5
	3.7	3.8							

a. Assuming that a simple linear regression model is appropriate, fit the regression model relating BOD (y) to the time (x). What is the estimate of σ^2?

b. What is the estimate of expected BOD level when the time is 15 days?

c. What change in mean BOD is expected when the time changes by 3 days?

d. Suppose that the time used is 6 days. Calculate the fitted value of y and the corresponding residual.

e. Calculate the fitted $\hat{y}_i$ for each value of x_i used to fit the model. Then construct a graph of $\hat{y}_i$ versus the corresponding observed values y_i and comment on what this plot would look like if the relationship between y and x was a deterministic (no random error) straight line. Does the plot actually obtained indicate that time is an effective regressor variable in predicting BOD?

11.2.10 Show that in a simple linear regression model the point $(\bar{x}, \bar{y})$ lies exactly on the least squares regression line.

Exercises for Section 11.4

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

11.4.1 WP Recall the regression of percent body fat on BMI from Exercise 11.2.1.

a. Estimate the error standard deviation.

b. Estimate the standard deviation of the slope.

c. What is the value of the t-statistic for the slope?

d. Test the hypothesis that $\beta_1 = 0$ at $\alpha = 0.05$. What is the P-value for this test?

11.4.2 Consider the simple linear regression model $y = 10 + 25x + \epsilon$ where the random error term is normally and independently distributed with mean zero and standard deviation 2. Use software to generate a sample of eight observations, one each at the levels $x = 10$, 12, 14, 16, 18, 20, 22, and 24.

a. Fit the linear regression model by least squares and find the estimates of the slope and intercept.

b. Find the estimate of σ^2.

c. Find the standard errors of the slope and intercept.

d. Now use software to generate a sample of 16 observations, two each at the same levels of x used previously. Fit the model using least squares.

e. Find the estimate of σ^2 for the new model in part (d). Compare this to the estimate obtained in part (b). What impact has the increase in sample size had on the estimate?

f. Find the standard errors of the slope and intercept using the new model from part (d). Compare these standard errors to the ones that you found in part (c). What impact has the increase in sample size had on the estimated standard errors?

11.4.3 WP Consider the following computer output.

```
The regression equation is
Y = 12.9 + 2.34 x
Predictor          Coef    SE Coef     T       P
Constant         12.857      1.032     ?       ?
X                 2.3445     0.1150    ?       ?
S = 1.48111  R-sq = 98.1%  R-sq(adj) = 97.9%

Analysis of Variance
Source             DF        SS       MS     F  P
Regression          1      912.43   912.43  ?  ?
Residual error      8       17.55      ?
Total               9      929.98
```

a. Fill in the missing information. You may use bounds for the P-values.

b. Can you conclude that the model defines a useful linear relationship?

c. What is your estimate of σ^2?

11.4.4 Consider the data from Exercise 11.2.2 on y = sales price and x = taxes paid.

a. Test H_0: $\beta_1 = 0$ using the t-test; use $\alpha = 0.05$.

b. Test H_0: $\beta_1 = 0$ using the analysis of variance with $\alpha = 0.05$. Discuss the relationship of this test to the test from part (a).

c. Estimate the standard errors of the slope and intercept.

d. Test the hypothesis that $\beta_0 = 0$.

11.4.5 Consider the data from Exercise 11.2.3 on x = compressive strength and y = intrinsic permeability of concrete.

a. Test for significance of regression using $\alpha = 0.05$. Find the P-value for this test. Can you conclude that the model specifies a useful linear relationship between these two variables?

b. Estimate σ^2 and the standard deviation of $\hat{\beta}_1$.

c. What is the standard error of the intercept in this model?

11.4.6 Consider the data from Exercise 11.2.4 on y = blood pressure rise and x = sound pressure level.

a. Test for significance of regression using $\alpha = 0.05$. What is the P-value for this test?

b. Estimate the standard errors of the slope and intercept.

c. Test H_0: $\beta_0 = 0$ versus H_1: $\beta_0 \neq 0$ using $\alpha = 0.05$. Find the P-value for this test.

11.4.7 Consider the data from Exercise 11.2.5 on y = green liquor Na_2S concentration and x = production in a paper mill.

a. Test for significance of regression using $\alpha = 0.05$. Find the P-value for this test.

b. Estimate the standard errors of the slope and intercept.

c. Test H_0: $\beta_0 = 0$ versus H_1: $\beta_0 \neq 0$ using $\alpha = 0.05$. What is the P-value for this test?

11.4.8 Consider the data from Exercise 11.2.6 on y = chloride concentration in surface streams and x = roadway area.

a. Test the hypothesis H_0: $\beta_1 = 0$ versus H_1: $\beta_1 \neq 0$ using the analysis of variance procedure with $\alpha = 0.01$.

b. Find the P-value for the test in part (a).

c. Estimate the standard errors of $\hat{\beta}_1$ and $\hat{\beta}_0$.

d. Test H_0: $\beta_1 = 0$ versus H_1: $\beta_0 \neq 0$ using $\alpha = 0.01$. What conclusions can you draw? Does it seem that the model might be a better fit to the data if the intercept were removed?

11.4.9 WP Consider the data in Exercise 11.2.9 on y = oxygen demand and x = time.

a. Test for significance of regression using $\alpha = 0.01$. Find the P-value for this test. What conclusions can you draw?

b. Estimate the standard errors of the slope and intercept.

c. Test the hypothesis that $\beta_0 = 0$.

11.4.10 WP SS Consider the data in Exercise 11.2.8 on y = deflection and x = stress level.

a. Test for significance of regression using $\alpha = 0.01$. What is the P-value for this test? State the conclusions that result from this test.

b. Does this model appear to be adequate?

c. Estimate the standard errors of the slope and intercept.

11.4.11 WP GO Tutorial An article in the *Journal of Clinical Endocrinology and Metabolism* ["Simultaneous and Continuous 24-Hour Plasma and Cerebrospinal Fluid Leptin Measurements: Dissociation of Concentrations in Central and Peripheral Compartments" (2004, Vol. 89, pp. 258–265)] reported on a study of the demographics of simultaneous and continuous 24-hour plasma and cerebrospinal fluid leptin measurements. The data follow:

y = BMI (kg/m^2):	19.92	20.59	29.02	20.78	25.97	
	20.39	23.29	17.27	35.24		
x = Age (yr):	45.5	34.6	40.6	32.9	28.2	30.1
	52.1	33.3	47.0			

a. Test for significance of regression using $\alpha = 0.05$. Find the P-value for this test. Can you conclude that the model specifies a useful linear relationship between these two variables?

b. Estimate σ^2 and the standard deviation of $\hat{\beta}_1$.

c. What is the standard error of the intercept in this model?

Exercises for Sections 11.5 and 11.6

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

11.6.1 WP Using the regression from Exercise 11.2.1,

a. Find a 95% confidence interval for the slope.

b. Find a 95% confidence interval for the mean percent body fat for a man with a BMI of 25.

c. Find a 95% prediction interval for the percent body fat for a man with a BMI of 25.

d. Which interval is wider, the confidence interval or the prediction interval? Explain briefly.

11.6.2 WP Refer to the data on y = house selling price and x = taxes paid in Exercise 11.2.2. Find a 95% confidence interval on each of the following:

a. β_1

b. β_0

c. Mean selling price when the taxes paid are $x = 7.50$

d. Compute the 95% prediction interval for selling price when the taxes paid are $x = 7.50$.

11.6.3 WP Refer to the data in Exercise 11.2.3 on y = intrinsic permeability of concrete and x = compressive strength. Find a 95% confidence interval on each of the following:

a. Slope

b. Intercept

c. Mean permeability when $x = 2.5$

d. Find a 95% prediction interval on permeability when $x = 2.5$. Explain why this interval is wider than the interval in part (c).

11.6.4 Exercise 11.2.4 presented data on y = blood pressure rise and x = sound pressure level. Find a 95% confidence interval on each of the following:

a. β_1

b. β_0

c. Mean blood pressure rise when the sound pressure level is 85 decibels

d. Find a 95% prediction interval on blood pressure rise when the sound pressure level is 85 decibels.

11.6.5 WP SS Consider the data in Exercise 11.2.5 on y = green liquor Na_2S concentration and x = production in a paper mill. Find a 99% confidence interval on each of the following:

a. β_1

b. β_0

c. Mean Na_2S concentration when production $x = 910$ tons/day

d. Find a 99% prediction interval on Na_2S concentration when $x = 910$ tons/day.

11.6.6 Exercise 11.2.6 presented data on chloride concentration y and roadway area x on watersheds in central Rhode Island. Find a 99% confidence interval on each of the following:

a. β_1

b. β_0

c. Mean chloride concentration when roadway area $x = 1.0\%$

d. Find a 99% prediction interval on chloride concentration when roadway area $x = 1.0\%$.

11.6.7 WP Refer to the data in Exercise 11.2.9 on oxygen demand. Find a 99% confidence interval on each of the following:

a. β_1

b. β_0

c. Find a 95% confidence interval on mean BOD when the time is 8 days.

11.6.8 Refer to the data in Exercise 11.2.7 on the microstructure of zirconia. Find a 95% confidence interval on each of the following:

a. Slope

b. Intercept

c. Mean length when $x = 1500$

d. Find a 95% prediction interval on length when $x = 1500$. Explain why this interval is wider than the interval in part (c).

Exercises for Section 11.7

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

11.7.1 WP SS Repeat Exercise 11.7.2 using an error term with a standard deviation of 4. What impact has increasing the error standard deviation had on the values of R^2?

11.7.2 Consider the simple linear regression model $y = 10 + 0.3x + \epsilon$ where the random error term is normally and independently distributed with mean zero and standard deviation 1. Use software to generate a sample of eight observations, one each at the levels $x = 10$, 12, 14, 16, 18, 20, 22, and 24.

a. Fit the linear regression model by least squares and find the estimates of the slope and intercept.

b. Find the estimate of σ^2.

c. Find the value of R^2.

d. Now use software to generate a new sample of eight observations, one each at the levels of $x = 10$, 14, 18, 22, 26, 30, 34, and 38. Fit the model using least squares.

e. Find R^2 for the new model in part (d). Compare this to the value obtained in part (c). What impact has the increase in the spread of the predictor variable x had on the value?

11.7.3 Refer to the compressive strength data in Exercise 11.2.3. Use the summary statistics provided to calculate R^2 and provide a practical interpretation of this quantity.

11.7.4 Refer to the data in Exercise 11.2.2 on house-selling price y and taxes paid x.

a. Find the residuals for the least squares model.

b. Prepare a normal probability plot of the residuals and interpret this display.

c. Plot the residuals versus $\hat{y}$ and versus x. Does the assumption of constant variance seem to be satisfied?

d. What proportion of total variability is explained by the regression model?

11.7.5 Refer to Exercise 11.2.6, which presented data on chloride concentration y and roadway area x.

a. What proportion of the total variability in chloride concentration is accounted for by the regression model?

b. Plot the residuals versus $\hat{y}$ and versus x. Interpret these plots.

c. Prepare a normal probability plot of the residuals. Does the normality assumption appear to be satisfied?

11.7.6 WP Refer to Exercise 11.2.4, which presented data on blood pressure rise y and sound pressure level x.

a. What proportion of total variability in blood pressure rise is accounted for by sound pressure level?

b. Prepare a normal probability plot of the residuals from this least squares model. Interpret this plot.

c. Plot residuals versus $\hat{y}$ and versus x. Comment on these plots.

11.7.7 WP VS SS An article in the *Journal of the American Statistical Association* ["Markov Chain Monte Carlo Methods for Computing Bayes Factors: A Comparative Review" (2001, Vol. 96, pp. 1122–1132)] analyzed the tabulated data on compressive strength parallel to the grain versus resin-adjusted density for specimens of radiata pine. The data are shown below.

Compressive Strength	Density	Compressive Strength	Density
3040	29.2	3840	30.7
2470	24.7	3800	32.7
3610	32.3	4600	32.6
3480	31.3	1900	22.1
3810	31.5	2530	25.3
2330	24.5	2920	30.8
1800	19.9	4990	38.9
3110	27.3	1670	22.1
3160	27.1	3310	29.2
2310	24.0	3450	30.1
4360	33.8	3600	31.4
1880	21.5	2850	26.7
3670	32.2	1590	22.1
1740	22.5	3770	30.3
2250	27.5	3850	32.0
2650	25.6	2480	23.2
4970	34.5	3570	30.3
2620	26.2	2620	29.9
2900	26.7	1890	20.8
1670	21.1	3030	33.2
2540	24.1	3030	28.2

a. Fit a regression model relating compressive strength to density.
b. Test for significance of regression with $\alpha = 0.05$.
c. Estimate σ^2 for this model.
d. Calculate R^2 for this model. Provide an interpretation of this quantity.
e. Prepare a normal probability plot of the residuals and interpret this display.
f. Plot the residuals versus $\hat{y}$ and versus x. Does the assumption of constant variance seem to be satisfied?

11.7.8 WP Consider the data in Exercise 11.2.5 on y = green liquor Na_2S concentration and x = paper machine production. Suppose that a 14th sample point is added to the original data where $y_{14} = 59$ and $x_{14} = 855$.

a. Prepare a scatter diagram of y versus x. Fit the simple linear regression model to all 14 observations.
b. Test for significance of regression with $\alpha = 0.05$.
c. Estimate σ^2 for this model.
d. Compare the estimate of σ^2 obtained in part (c) with the estimate of σ^2 obtained from the original 13 points. Which estimate is larger and why?
e. Compute the residuals for this model. Does the value of e_{14} appear unusual?
f. Prepare and interpret a normal probability plot of the residuals.
g. Plot the residuals versus $\hat{y}$ and versus x. Comment on these graphs.

11.7.9 Show that an equivalent way to define the test for significance of regression in simple linear regression is to base the test on R^2 as follows: to test H_0: $\beta_1 = 0$ versus H_0: $\beta_1 \neq 0$, calculate

$$F_0 = \frac{R^2(n-2)}{1-R^2}$$

and to reject H_0: $\beta_1 = 0$ if the computed value $f_0 > f_{\alpha,1,n-2}$. Suppose that a simple linear regression model has been fit to $n = 25$ observations and $R^2 = 0.90$.

a. Test for significance of regression at $\alpha = 0.05$.
b. What is the smallest value of R^2 that would lead to the conclusion of a significant regression if $\alpha = 0.05$?

11.7.10 WP **Studentized Residuals.** Show that the variance of the ith residual is

$$V(e_i) = \sigma^2\left[1 - \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}\right)\right]$$

Hint:

$$\text{cov}(Y_i, \hat{Y}_i) = \sigma^2\left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}\right].$$

The ith studentized residual is defined as

$$r_i = \frac{e_i}{\sqrt{\hat{\sigma}^2\left[1 - \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}\right)\right]}}$$

a. Explain why r_i has unit standard deviation.
b. Do the standardized residuals have unit standard deviation?
c. Discuss the behavior of the studentized residual when the sample value x_i is very close to the middle of the range of x.
d. Discuss the behavior of the studentized residual when the sample value x_i is very near one end of the range of x.

Exercises for Section 11.8

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

11.8.1 WP Suppose that data are obtained from 20 pairs of (x, y) and the sample correlation coefficient is 0.8.

a. Test the hypothesis that H_0: $\rho = 0$ against H_1: $\rho \neq 0$ with $\alpha = 0.05$. Calculate the P-value.
b. Test the hypothesis that H_1: $\rho = 0.5$ against H_1: $\rho \neq 0.5$ with $\alpha = 0.05$. Calculate the P-value.
c. Construct a 95% two-sided confidence interval for the correlation coefficient. Explain how the questions in parts (a) and (b) could be answered with a confidence interval.

11.8.2 A random sample of $n = 25$ observations was made on the time to failure of an electronic component and the temperature in the application environment in which the component was used.

a. Given that $\hat{\rho} = 0.83$, test the hypothesis that $\rho = 0$ using $\alpha = 0.05$. What is the P-value for this test?

b. Find a 95% confidence interval on ρ.

c. Test the hypothesis H_0: $\rho = 0.8$ versus H_1: $\rho \neq 0.8$, using $\alpha = 0.05$. Find the P-value for this test.

11.8.3 The data in the following table give x = the water content of snow on April 1 and y = the yield from April to July (in inches) on the Snake River watershed in Wyoming for 1919 to 1935. (The data were taken from an article in *Research Notes*, Vol. 61, 1950, Pacific Northwest Forest Range Experiment Station, Oregon.)

x	y	x	y
23.1	10.5	37.9	22.8
32.8	16.7	30.5	14.1
31.8	18.2	25.1	12.9
32.0	17.0	12.4	8.8
30.4	16.3	35.1	17.4
24.0	10.5	31.5	14.9
39.5	23.1	21.1	10.5
24.2	12.4	27.6	16.1
52.5	24.9		

a. Estimate the correlation between Y and X.

b. Test the hypothesis that $\rho = 0$ using $\alpha = 0.05$.

c. Fit a simple linear regression model and test for significance of regression using $\alpha = 0.05$. What conclusions can you draw? How is the test for significance of regression related to the test on ρ in part (b)?

d. Analyze the residuals and comment on term list.

11.8.4 The weight and systolic blood pressure of 26 randomly selected males in the age group 25 to 30 are shown below. Assume that weight and blood pressure are jointly normally distributed.

Subject	Weight	Systolic BP	Subject	Weight	Systolic BP
1	165	130	14	172	153
2	167	133	15	159	128
3	180	150	16	168	132
4	155	128	17	174	149
5	212	151	18	183	158
6	175	146	19	215	150
7	190	150	20	195	163
8	210	140	21	180	156
9	200	148	22	143	124
10	149	125	23	240	170
11	158	133	24	235	165
12	169	135	25	192	160
13	170	150	26	187	159

a. Find a regression line relating systolic blood pressure to weight.

b. Test for significance of regression using $\alpha = 0.05$.

c. Estimate the correlation coefficient.

d. Test the hypothesis that $\rho = 0$, using $\alpha = 0.05$.

e. Test the hypothesis that $\rho = 0$, using $\alpha = 0.05$.

f. Construct a 95% confidence interval for the correlation coefficient.

11.8.5 SS In an article in *IEEE Transactions on Instrumentation and Measurement* (2001, Vol. 50, pp. 986–990), researchers reported on a study of the effects of reducing current draw in a magnetic core by electronic means. They measured the current in a magnetic winding with and without the electronics in a paired experiment. Data for the case without electronics are provided in the following table.

Supply Voltage	Current Without Electronics (mA)
0.66	7.32
1.32	12.22
1.98	16.34
2.64	23.66
3.3	28.06
3.96	33.39
4.62	34.12
3.28	39.21
5.94	44.21
6.6	47.48

a. Graph the data and fit a regression line to predict current without electronics to supply voltage. Is there a significant regression at $\alpha = 0.05$? What is the P-value?

b. Estimate the correlation coefficient.

c. Test the hypothesis that $\rho = 0$ against the alternative $\rho \neq 0$ with $\alpha = 0.05$. What is the P-value?

d. Compute a 95% confidence interval for the correlation coefficient.

11.8.6 The monthly absolute estimate of global (land and ocean combined) temperature indexes (degrees C) in 2000 and 2001 (*www.ncdc.noaa.gov/oa/climate/*) are:

2000: 12.28, 12.63, 13.22, 14.21, 15.13, 15.82, 16.05, 16.02, 15.29, 14.29, 13.16, 12.47

2001: 12.44, 12.55, 13.35, 14.22, 15.28, 15.99, 16.23, 16.17, 15.44, 14.52, 13.52, 12.61

a. Graph the data and fit a regression line to predict 2001 temperatures from those in 2000. Is there a significant regression at $\alpha = 0.05$? What is the P-value?

b. Estimate the correlation coefficient.

c. Test the hypothesis that $\rho = 0.9$ against the alternative $\rho \neq 0.9$ with $\alpha = 0.05$. What is the P-value?

d. Compute a 95% confidence interval for the correlation coefficient.

Exercises for Section 11.9

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

11.9.1 **SS** Determine if the following models are intrinsically linear. If yes, determine the appropriate transformation to generate the linear model.

a. $Y = \beta_0 x^{\beta_1}\ \varepsilon$

b. $Y = \dfrac{3 + 5x}{x} + \varepsilon$

c. $Y = \beta_0 \beta_1^x\ \varepsilon$

d. $Y = \dfrac{x}{\beta_0 x + \beta_1 + x\varepsilon}$

11.9.2 **VS** The vapor pressure of water at various temperatures is in the table below:

a. Draw a scatter diagram of these data. What type of relationship seems appropriate in relating y to x?
b. Fit a simple linear regression model to these data.
c. Test for significance of regression using $\alpha = 0.05$. What conclusions can you draw?
d. Plot the residuals from the simple linear regression model versus $\hat{y}_i$. What do you conclude about model adequacy?
e. The Clausius–Clapeyron relationship states that $\ln(P_v) \propto -\frac{1}{T}$ where P_v is the vapor pressure of water. Repeat parts (a)–(d) using an appropriate transformation.

Observation Number	Temperature (*K*)	Vapor pressure (mm Hg)
1	273	4.6
2	283	9.2
3	293	17.5
4	303	31.8
5	313	55.3
6	323	92.5
7	333	149.4
8	343	233.7
9	353	355.1
10	363	525.8
11	373	760.0

11.9.3 An electric utility is interested in developing a model relating peak-hour demand (y in kilowatts) to total monthly energy usage during the month (x, in kilowatt hours). Data for 50 residential customers are shown in the table below.

a. Draw a scatter diagram of y versus x.
b. Fit the simple linear regression model.
c. Test for significance of regression using $\alpha = 0.05$.
d. Plot the residuals versus $\hat{y}_i$ and comment on the underlying regression assumptions. Specifically, does it seem that the equality of variance assumption is satisfied?
e. Find a simple linear regression model using $\sqrt{y}$ as the response. Does this transformation on y stabilize the inequality of variance problem noted in part (d)?

Customer	*x*	*y*	Customer	*x*	*y*
1	679	0.79	26	1434	0.31
2	292	0.44	27	837	4.20
3	1012	0.56	28	1748	4.88
4	493	0.79	29	1381	3.48
5	582	2.70	30	1428	7.58
6	1156	3.64	31	1255	2.63
7	997	4.73	32	1777	4.99
8	2189	9.50	33	370	0.59
9	1097	5.34	34	2316	8.19
10	2078	6.85	35	1130	4.79
11	1818	5.84	36	463	0.51
12	1700	5.21	37	770	1.74
13	747	3.25	38	724	4.10
14	2030	4.43	39	808	3.94
15	1643	3.16	40	790	0.96
16	414	0.50	41	783	3.29
17	354	0.17	42	406	0.44
18	1276	1.88	43	1242	3.24
19	745	0.77	44	658	2.14
20	795	3.70	45	1746	5.71
21	540	0.56	46	895	4.12
22	874	1.56	47	1114	1.90
23	1543	5.28	48	413	0.51
24	1029	0.64	49	1787	8.33
25	710	4.00	50	3560	14.94

Exercises for Section 11.10

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutoring Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

11.10.1 WP A study was conducted attempting to relate home ownership to family income. Twenty households were selected and family income was estimated along with information concerning home ownership ($y = 1$ indicates *yes* and $y = 0$ indicates *no*). The data are shown in the table below.

Household	Income	Home Ownership Status
1	38,000	0
2	51,200	1
3	39,600	0
4	43,400	1
5	47,700	0
6	53,000	0
7	41,500	1
8	40,800	0
9	45,400	1
10	52,400	1
11	38,700	1
12	40,100	0
13	49,500	1
14	38,000	0
15	42,000	1
16	54,000	1
17	51,700	1
18	39,400	0
19	40,900	0
20	52,800	1

a. Fit a logistic regression model to the response variable y. Use a simple linear regression model as the structure for the linear predictor.
b. Is the logistic regression model in part (a) adequate?
c. Provide an interpretation of the parameter β_1 in this model.

11.10.2 WP The compressive strength of an alloy fastener used in aircraft construction is being studied. Ten loads were selected over the range 2500–4300 psi, and a number of fasteners were tested at those loads. The numbers of fasteners failing at each load were recorded. The complete test data are shown in the following table.

Load, x (psi)	Sample Size, n	Number Failing, r
2500	50	10
2700	70	17
2900	100	30
3100	60	21
3300	40	18
3500	85	43
3700	90	54
3900	50	33
4100	80	60
4300	65	51

a. Fit a logistic regression model to the data. Use a simple linear regression model as the structure for the linear predictor.
b. Is the logistic regression model in part (a) adequate?

11.10.3 WP SS The market research department of a soft drink manufacturer is investigating the effectiveness of a price discount coupon on the purchase of a 2-liter beverage product. A sample of 5500 customers was given coupons for varying price discounts between 5 and 25 cents. The response variable was the number of coupons in each price discount category redeemed after one month. The data follow.

Discount, x	Sample Size, n	Number Redeemed, r
5	500	100
7	500	122
9	500	147
11	500	176
13	500	211
15	500	244
17	500	277
19	500	310
21	500	343
23	500	372
25	500	391

a. Fit a logistic regression model to the data. Use a simple linear regression model as the structure for the linear predictor.
b. Is the logistic regression model in part (a) adequate?
c. Draw a graph of the data and the fitted logistic regression model.

d. Expand the linear predictor to include a quadratic term. Is there any evidence that this quadratic term is required in the model?

e. Draw a graph of this new model on the same plot that you prepared in part (c). Does the expanded model visually provide a better fit to the data than the original model from part (a)?

11.10.4 WP A study was performed to investigate new automobile purchases. A sample of 20 families was selected. Each family was surveyed to determine the age of their oldest vehicle and their total family income. A follow-up survey was conducted 6 months later to determine if they had actually purchased a new vehicle during that time period ($y = 1$ indicates *yes* and $y = 0$ indicates *no*). The data from this study are shown in the table below.

Income, x_1	Age, x_2	y	Income, x_1	Age, x_2	y
45,000	2	0	37,000	5	1
40,000	4	0	31,000	7	1
60,000	3	1	40,000	4	1
50,000	2	1	75,000	2	0
55,000	2	0	43,000	9	1
50,000	5	1	49,000	2	0
35,000	7	1	37,500	4	1
65,000	2	1	71,000	1	0
53,000	2	0	34,000	5	0
48,000	1	0	27,000	6	0

a. Fit a logistic regression model to the data.

b. Is the logistic regression model in part (a) adequate?

c. Interpret the model coefficients β_1 and β_2.

d. What is the estimated probability that a family with an income of $45,000 and a car that is 5 years old will purchase a new vehicle in the next 6 months?

e. Expand the linear predictor to include an interaction term. Is there any evidence that this term is required in the model?

11.10.5 WP The World Health Organization defines *obesity* in adults as having a body mass index (BMI) higher than 30. Of the 250 men in the study mentioned in Exercise 11.2.1, 23 are by this definition obese. How good is waist (size in inches) as a predictor of obesity? A logistic regression model was fit to the data:

$$\log\left(\frac{p}{1-p}\right) = -41.828 + 0.9864 \text{ waist}$$

where p is the probability of being classified as obese.

a. Does the probability of being classified as obese increase or decrease as a function of waist size? Explain.

b. What is the estimated probability of being classified as obese for a man with a waist size of 36 inches?

c. What is the estimated probability of being classified as obese for a man with a waist size of 42 inches?

d. What is the estimated probability of being classified as obese for a man with a waist size of 48 inches?

e. Make a plot of the estimated probability of being classified as obese as a function of waist size.

Supplemental Exercises for Chapter 11

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

11.S7 WP An article in the *IEEE Transactions on Instrumentation and Measurement* ["Direct, Fast, and Accurate Measurement of V_T and K of MOS Transistor Using V_T-Sift Circuit" (1991, Vol. 40, pp. 951–955)] described the use of a simple linear regression model to express drain current y (in milliamperes) as a function of ground-to-source voltage x (in volts). The data are as follows:

y	x	y	x
0.734	1.1	1.50	1.6
0.886	1.2	1.66	1.7
1.04	1.3	1.81	1.8
1.19	1.4	1.97	1.9
1.35	1.5	2.12	2.0

a. Draw a scatter diagram of these data. Does a straight-line relationship seem plausible?

b. Fit a simple linear regression model to these data.

c. Test for significance of regression using $\alpha = 0.05$. What is the P-value for this test?

d. Find a 95% confidence interval estimate on the slope.

e. Test the hypothesis $H_0: \beta_0 = 0$ versus $H_1: \beta_0 \neq 0$ using $\alpha = 0.05$. What conclusions can you draw?

11.S8 Show that, for the simple linear regression model, the following statements are true:

a. $\sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$

b. $\sum_{i=1}(y_i - \hat{y}_i)x_i = 0$

c. $\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i = \overline{y}$

11.S9 SS Consider the following data. Suppose that the relationship between Y and x is hypothesized to be $Y = (\beta_0 + \beta_1 x + \epsilon)^{-1}$.

Fit an appropriate model to the data. Does the assumed model form seem reasonable?

x	10	15	18	12	9	8	11	6
y	0.1	0.13	0.09	0.15	0.20	0.21	0.18	0.24

11.S10 WP The strength of paper used in the manufacture of cardboard boxes (y) is related to the percentage of hardwood concentration in the original pulp (x). Under controlled conditions, a pilot plant manufactures 16 samples, each from a different batch of pulp, and measures the tensile strength. The data follow:

y	101.4	117.4	117.1	106.2
x	1.0	1.5	1.5	1.5
y	131.9	146.9	146.8	133.9
x	2.0	2.0	2.2	2.4
y	111.0	123.0	125.1	145.2
x	2.5	2.5	2.8	2.8
y	134.3	144.5	143.7	146.9
x	3.0	3.0	3.2	3.3

a. Fit a simple linear regression model to the data.
b. Test for significance of regression using $\alpha = 0.05$.
c. Construct a 90% confidence interval on the slope β_1.
d. Construct a 90% confidence interval on the intercept β_0.
e. Construct a 95% confidence interval on the mean strength at $x = 2.5$.
f. Analyze the residuals and comment on model adequacy.

11.S11 WP The data in the table below, adapted from Montgomery, Peck, and Vining (2012), present the number of certified mental defectives per 10,000 of estimated population in the United Kingdom (y) and the number of radio receiver licenses issued (x) by the BBC (in millions) for the years 1924 through 1937. Fit a regression model relating y and x. Comment on the model. Specifically, does the existence of a strong correlation imply a cause-and-effect relationship?

Year	y	x	Year	y	x
1924	8	1.350	1931	16	4.620
1925	8	1.960	1932	18	5.497
1926	9	2.270	1933	19	6.260
1927	10	2.483	1934	20	7.012
1928	11	2.730	1935	21	7.618
1929	11	3.091	1936	22	8.131
1930	12	3.674	1937	23	8.593

11.S12 An article in *Air and Waste* ["Update on Ozone Trends in California's South Coast Air Basin" (1993, Vol. 43, pp. 226–240)] reported on a study of the ozone levels on the South Coast air basin of California for the years 1976–1991. The author believes that the number of days that the ozone level exceeds 0.20 parts per million depends on the seasonal meteorological index (the seasonal average 850 millibar temperature). The data follow:

Year	Days	Index	Year	Days	Index
1976	91	16.7	1984	81	18.0
1977	105	17.1	1985	65	17.2
1978	106	18.2	1986	61	16.9
1979	108	18.1	1987	48	17.1
1980	88	17.2	1988	61	18.2
1981	91	18.2	1989	43	17.3
1982	58	16.0	1990	33	17.5
1983	82	17.2	1991	36	16.6

a. Construct a scatter diagram of the data.
b. Fit a simple linear regression model to the data. Test for significance of regression.
c. Find a 95% CI on the slope β_1.
d. Analyze the residuals and comment on model adequacy.

11.S13 An article in the *Journal of Applied Polymer Science* ["Preparation and Characterizations of Thermotropic Copolyesters of *p*-Hydroxybenzoic Acid, Sebacic Acid, and Hydroquinone" (1995, Vol. 56, pp. 471–476)] reported on a study of the effect of the mole ratio of sebacic acid on the intrinsic viscosity of copolyesters. The data follow:

Mole ratio x	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3
Viscosity y	0.45	0.20	0.34	0.58	0.70	0.57	0.55	0.44

a. Construct a scatter diagram of the data.
b. Fit a simple linear repression model.
c. Test for significance of regression. Calculate R^2 for the model.
d. Analyze the residuals and comment on model adequacy.

11.S14 Two different methods can be used for measuring the temperature of the solution in a Hall cell used in aluminum smelting, a thermocouple implanted in the cell and an indirect measurement produced from an IR device. The indirect method is preferable because the thermocouples are eventually destroyed by the solution. Consider the following 10 measurements:

Thermocouple	921	935	916	920	940
IR	918	934	924	921	945
Thermocouple	936	925	940	933	927
IR	930	919	943	932	935

a. Construct a scatter diagram for these data, letting x = thermocouple measurement and y = IR measurement.

b. Fit a simple linear regression model.
c. Test for significance of a regression and calculate R^2. What conclusions can you draw?
d. Is there evidence to support a claim that both devices produce equivalent temperature measurements? Formulate and test an appropriate hypothesis to support this claim.
e. Analyze the residuals and comment on model adequacy.

11.S15 The data below related diamond carats to purchase prices. It appeared in Singapore's *Business Times*, February 18, 2000.

Carat	Price	Carat	Price
0.3	1302	0.33	1327
0.3	1510	0.33	1098
0.3	1510	0.34	1693
0.3	1260	0.34	1551
0.31	1641	0.34	1410
0.31	1555	0.34	1269
0.31	1427	0.34	1316
0.31	1427	0.34	1222
0.31	1126	0.35	1738
0.31	1126	0.35	1593
0.32	1468	0.35	1447
0.32	1202	0.35	1255
0.36	1635	0.45	1572
0.36	1485	0.46	2942
0.37	1420	0.48	2532
0.37	1420	0.5	3501
0.4	1911	0.5	3501
0.4	1525	0.5	3501
0.41	1956	0.5	3293
0.43	1747	0.5	3016

a. Graph the data. What is the relation between carat and price? Is there an outlier?
b. What would you say to the person who purchased the diamond that was an outlier?
c. Fit two regression models, one with all the data and the other with unusual data omitted. Estimate the slope coefficient with a 95% confidence interval in both cases. Comment on any difference.

Chapter 12 Exercises

Exercises for Section 12.1

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

12.1.1 Exercise 11.2.1 described a regression model between percent of body fat (%BF) as measured by immersion and BMI from a study on 250 male subjects. The researchers also measured 13 physical characteristics of each man, including his age (yrs), height (in), and waist size (in).

A regression of percent of body fat with both height and waist as predictors shows the following computer output:

```
             Estimate Std.  Error   t-value   Pr(>|t|)
(Intercept)  −3.10088    7.68611   −0.403    0.687
Height       −0.60154    0.10994   −5.472    1.09e−07
Waist         1.77309    0.07158   24.770    < 2e −16
Residual standard error: 4.46 on 247 degrees of freedom
Multiple R-squared: 0.7132, Adjusted R-squared: 0.7109
F-statistic: 307.1 on 2 and 247 DF, p-value: < 2.2e-16
```

a. Write out the regression model if

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 2.9705 & -4.0042E{-}2 & -4.1679E{-}2 \\ -0.04004 & 6.0774E{-}4 & -7.3875E{-}5 \\ -4.1679E{-}3 & -7.3875E{-}5 & 2.5766E{-}4 \end{bmatrix}$$

and

$$(\mathbf{X}'\mathbf{y}) = \begin{bmatrix} 4757.9 \\ 334335.8 \\ 179706.7 \end{bmatrix}$$

b. Verify that the model found from technology is correct to at least two decimal places.
c. What is the predicted body fat of a man who is 6-ft tall with a 34-in. waist?

12.1.2 Hsuie, Ma, and Tsai ["Separation and Characterizations of Thermotropic Copolyesters of p-Hydroxybenzoic Acid, Sebacic Acid, and Hydroquinone" *Journal of Applied Polymer Science* (1995, Vol. 56, pp. 471–476)] studied the effect of the molar ratio of sebacic acid (the regressor) on the intrinsic viscosity of copolyesters (the response). The following display presents the data.

Ratio	Viscosity
1.0	0.45
0.9	0.20
0.8	0.34

(*continued*)

0.7	0.58
0.6	0.70
0.5	0.57
0.4	0.55
0.3	0.44

a. Construct a scatter plot of the data.

b. Fit a second-order prediction equation.

12.1.3 Can the percentage of the workforce who are engineers in each U.S. state be predicted by the amount of money spent on higher education (as a percent of gross domestic product), on venture capital (dollars per $1000 of gross domestic product) for high-tech business ideas, and state funding (in dollars per student) for major research universities? Data for all 50 states and a software package revealed the following results:

```
                Estimate   Std. Error t value Pr(>|t|)
(Intercept)    1.051e+00  1.567e-01   6.708  2.5e-08 ***
Venture cap    9.514e-02  3.910e-02   2.433  0.0189 *
State funding  4.106e-06  1.437e-05   0.286  0.7763
Higher.eD     -1.673e-01  2.595e-01  -0.645  0.5223
Residual standard error: 0.3007 on 46 degrees of
freedom
Multiple R-squared: 0.1622, Adjusted R-squared: 0.1075
F-statistic: 2.968 on 3 and 46 DF, p-value: 0.04157
```

a. Write the equation predicting the percent of engineers in the workforce.

b. For a state that has $1 per $1000 in venture capital, spends $10,000 per student on funding for major research universities, and spends 0.5% of its GDP on higher education, what percent of engineers do you expect to see in the workforce?

c. If the state in part (b) actually had 1.5% engineers in the workforce, what would the residual be?

12.1.4 WP SS A regression model is to be developed for predicting the ability of soil to absorb chemical contaminants. Ten observations have been taken on a soil absorption index (y) and two regressors: x_1 = amount of extractable iron ore and x_2 = amount of bauxite. We wish to fit the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$. Some necessary quantities are:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1.17991 & -7.30982\text{ E-}3 & 7.3006\text{ E-}4 \\ -7.30982\text{ E-}3 & 7.9799\text{ E-}5 & -1.23713\text{ E-}4 \\ 7.3006\text{ E-}4 & -1.23713\text{ E-}4 & 4.6576\text{ E-}4 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 220 \\ 36{,}768 \\ 9{,}965 \end{bmatrix}$$

a. Estimate the regression coefficients in the model specified.

b. What is the predicted value of the absorption index y when $x_1 = 200$ and $x_2 = 50$?

12.1.5 The data from a patient satisfaction survey in a hospital are shown next.

Observation	Age	Severity	Surg-Med	Anxiety	Satisfaction
1	55	50	0	2.1	68
2	46	24	1	2.8	77
3	30	46	1	3.3	96
4	35	48	1	4.5	80
5	59	58	0	2.0	43
6	61	60	0	5.1	44
7	74	65	1	5.5	26
8	38	42	1	3.2	88
9	27	42	0	3.1	75
10	51	50	1	2.4	57
11	53	38	1	2.2	56
12	41	30	0	2.1	88
13	37	31	0	1.9	88
14	24	34	0	3.1	102
15	42	30	0	3.0	88
16	50	48	1	4.2	70
17	58	61	1	4.6	52
18	60	71	1	5.3	43
19	62	62	0	7.2	46
20	68	38	0	7.8	56
21	70	41	1	7.0	59
22	79	66	1	6.2	26
23	63	31	1	4.1	52
24	39	42	0	3.5	83
25	49	40	1	2.1	75

The regressor variables are the patient's age, an illness severity index (higher values indicate greater severity), an indicator variable denoting whether the patient is a medical patient (0) or a surgical patient (1), and an anxiety index (higher values indicate greater anxiety).

a. Fit a multiple linear regression model to the satisfaction response using age, illness severity, and the anxiety index as the regressors.

b. Estimate σ^2.

c. Find the standard errors of the regression coefficients.

d. Are all of the model parameters estimated with nearly the same precision? Why or why not?

12.1.6 The electric power consumed each month by a chemical plant is thought to be related to the average ambient temperature (x_1), the number of days in the month (x_2), the average product purity (x_3), and the tons of product produced (x_4). The past year's historical data are available and are presented below.

y	x_1	x_2	x_3	x_4
240	25	24	91	100
236	31	21	90	95
270	45	24	88	110
274	60	25	87	88
301	65	25	91	94
316	72	26	94	99
300	80	25	87	97
296	84	25	86	96
267	75	24	88	110
276	60	25	91	105
288	50	25	90	100
261	38	23	89	98

a. Fit a multiple linear regression model to these data.
b. Estimate σ^2.
c. Compute the standard errors of the regression coefficients. Are all of the model parameters estimated with the same precision? Why or why not?
d. Predict power consumption for a month in which $x_1 = 75°\text{F}$, $x_2 = 24$ days, $x_3 = 90\%$, and $x_4 = 98$ tons.

12.1.7 An article in *Electronic Packaging and Production* (2002, Vol. 42) considered the effect of X-ray inspection of integrated circuits. The rads (radiation dose) were studied as a function of current (in milliamps) and exposure time (in minutes). The data are shown below.

Rads	mAmps	Exposure Time
7.4	10	0.25
14.8	10	0.5
29.6	10	1
59.2	10	2
88.8	10	3
296	10	10
444	10	15
592	10	20
11.1	15	0.25
22.2	15	0.5
44.4	15	1
88.8	15	2
133.2	15	3
444	15	10
666	15	15
888	15	20
14.8	20	0.25
29.6	20	0.5
59.2	20	1
118.4	20	2
177.6	20	3
592	20	10
888	20	15
1184	20	20
22.2	30	0.25
44.4	30	0.5
88.8	30	1
177.6	30	2
266.4	30	3
888	30	10
1332	30	15
1776	30	20
29.6	40	0.25
59.2	40	0.5
118.4	40	1
236.8	40	2
355.2	40	3
1184	40	10
1776	40	15
2368	40	20

a. Fit a multiple linear regression model to these data with rads as the response.
b. Estimate σ^2 and the standard errors of the regression coefficients.
c. Use the model to predict rads when the current is 15 milliamps and the exposure time is 5 seconds.

12.1.8 WP SS The pull strength of a wire bond is an important characteristic. The table below gives information on pull strength (y), die height (x_1), post height (x_2), loop height (x_3), wire length (x_4), bond width on the die (x_5), and bond width on the post (x_6).

y	x_1	x_2	x_3	x_4	x_5	x_6
8.0	5.2	19.6	29.6	94.9	2.1	2.3
8.3	5.2	19.8	32.4	89.7	2.1	1.8
8.5	5.8	19.6	31.0	96.2	2.0	2.0
8.8	6.4	19.4	32.4	95.6	2.2	2.1
9.0	5.8	18.6	28.6	86.5	2.0	1.8
9.3	5.2	18.8	30.6	84.5	2.1	2.1
9.3	5.6	20.4	32.4	88.8	2.2	1.9
9.5	6.0	19.0	32.6	85.7	2.1	1.9
9.8	5.2	20.8	32.2	93.6	2.3	2.1
10.0	5.8	19.9	31.8	86.0	2.1	1.8
10.3	6.4	18.0	32.6	87.1	2.0	1.6
10.5	6.0	20.6	33.4	93.1	2.1	2.1

(*continued*)

y	x_1	x_2	x_3	x_4	x_5	x_6
10.8	6.2	20.2	31.8	83.4	2.2	2.1
11.0	6.2	20.2	32.4	94.5	2.1	1.9
11.3	6.2	19.2	31.4	83.4	1.9	1.8
11.5	5.6	17.0	33.2	85.2	2.1	2.1
11.8	6.0	19.8	35.4	84.1	2.0	1.8
12.3	5.8	18.8	34.0	86.9	2.1	1.8
12.5	5.6	18.6	34.2	83.0	1.9	2.0

a. Fit a multiple linear regression model using x_2, x_3, x_4, and x_5 as the regressors.
b. Estimate σ^2.
c. Find the $se(\hat{\beta}_j)$. How precisely are the regression coefficients estimated, in your opinion?
d. Use the model from part (a) to predict pull strength when $x_2 = 20$, $x_3 = 30$, $x_4 = 90$, and $x_5 = 2.0$.

12.1.9 An article in *IEEE Transactions on Instrumentation and Measurement* ["Measurement and Calculation of Powered Mixture Permittivities" (2001, Vol. 50, pp. 1066–1070)] reported on a study that had analyzed powdered mixtures of coal and limestone for permittivity. The errors in the density measurement was the response. The data are reported in the following table.

Density	Dielectric Constant	Loss Factor
0.749	2.05	0.016
0.798	2.15	0.02
0.849	2.25	0.022
0.877	2.3	0.023
0.929	2.4	0.026
0.963	2.47	0.028
0.997	2.54	0.031
1.046	2.64	0.034
1.133	2.85	0.039
1.17	2.94	0.042
1.215	3.05	0.045

a. Fit a multiple linear regression model to these data with the density as the response.
b. Estimate σ^2 and the standard errors of the regression coefficients.
c. Use the model to predict the density when the dielectric constant is 2.5 and the loss factor is 0.03.

12.1.10 An article in *Biotechnology Progress* ["Optimization of Conditions for Bacteriocin Extraction in PEG/Salt Aqueous Two-Phase Systems Using Statistical Experimental Designs" (2001, Vol. 17, pp. 366–368)] reported on an experiment to investigate and optimize nisin extraction in aqueous two-phase systems (ATPS). The nisin recovery was the dependent variable (y). The two regressor variables were concentration (%) of PEG 4000 (denoted as x_1) and concentration (%) of Na_2SO_4 (denoted as x_2). The data are shown below.

x_1	x_2	y
13	11	62.8739
15	11	76.1328
13	13	87.4667
15	13	102.3236
14	12	76.1872
14	12	77.5287
14	12	76.7824
14	12	77.4381
14	12	78.7417

a. Fit a multiple linear regression model to these data.
b. Estimate σ^2 and the standard errors of the regression coefficients.
c. Use the model to predict the nisin recovery when $x_1 = 14.5$ and $x_2 = 12.5$.

12.1.11 SS VS An article in *Optical Engineering* ["Operating Curve Extraction of a Correlator's Filter" (2004, Vol. 43, pp. 2775–2779)] reported on the use of an optical correlator to perform an experiment by varying brightness and contrast. The resulting modulation is characterized by the useful range of gray levels. The data follow:

Brightness (%):	54	61	65	100	100	100	50	57	54
Contrast (%):	56	80	70	50	65	80	25	35	26
Useful range (ng):	96	50	50	112	96	80	155	144	255

a. Fit a multiple linear regression model to these data.
b. Estimate σ^2.
c. Compute the standard errors of the regression coefficients.
d. Predict the useful range when brightness = 80 and contrast = 75.

12.1.12 The table below and on the next page presents statistics for the National Hockey League teams from the 2008–2009 season (*The Sports Network*). Fit a multiple linear regression model that relates *wins* to the variables *GF* through *FG*. Teams play 82 games, $W = 82 - L - T - OTL$, but such a model does not help build a better team. Estimate σ^2 and find the standard errors of the regression coefficients for your model.

12.1.13 A study was performed on wear of a bearing and its relationship to x_1 = oil viscosity and x_2 = load. The following data were obtained.

y	x_1	x_2
293	1.6	851
230	15.5	816
172	22.0	1058
91	43.0	1201
113	33.0	1357
125	40.0	1115

a. Fit a multiple linear regression model to these data.
b. Estimate σ^2 and the standard errors of the regression coefficients.
c. Use the model to predict wear when $x_1 = 25$ and $x_2 = 1000$.
d. Fit a multiple linear regression model with an interaction term to these data.
e. Estimate σ^2 and $se(\hat{\beta}_j)$ for this new model. How did these quantities change? Does this tell you anything about the value of adding the interaction term to the model?
f. Use the model in part (d) to predict when $x_1 = 25$ and $x_2 = 1000$. Compare this prediction with the predicted value from part (c).

Team	W	L	OTL	PTS	GF	GA	ADV	PPGF	PCTG	PEN	BMI	AVG	SHT	PPGA	PKPCT	SHGF	SHGA	FG
Anaheim	42	33	7	91	238	235	309	73	23.6	1418	8	17.4	385	78	79.7	6	6	43
Atlanta	35	41	6	76	250	279	357	69	19.3	1244	12	15.3	366	88	76	13	9	39
Boston	53	19	10	116	270	190	313	74	23.6	1016	12	12.5	306	54	82.4	8	7	47
Buffalo	41	32	9	91	242	229	358	75	21	1105	16	13.7	336	61	81.8	7	4	44
Carolina	45	30	7	97	236	221	374	70	18.7	786	16	9.8	301	59	80.4	8	7	39
Columbus	41	31	10	92	220	223	322	41	12.7	1207	20	15	346	62	82.1	8	9	41
Calgary	46	30	6	98	251	246	358	61	17	1281	18	15.8	349	58	83.4	6	13	37
Chicago	46	24	12	104	260	209	363	70	19.3	1129	28	14.1	330	64	80.6	10	5	43
Colorado	32	45	5	69	190	253	318	50	15.7	1044	18	13	318	64	79.9	4	5	31
Dallas	36	35	11	83	224	251	351	54	15.4	1134	10	14	327	70	78.6	2	2	38
Detroit	51	21	10	112	289	240	353	90	25.5	810	14	10	327	71	78.3	6	4	46
Edmonton	38	35	9	85	228	244	354	60	17	1227	20	15.2	338	76	77.5	3	8	39
Florida	41	30	11	93	231	223	308	51	16.6	884	16	11	311	54	82.6	7	6	39
Los Angeles	34	37	11	79	202	226	360	69	19.2	1191	16	14.7	362	62	82.9	4	7	39
Minnesota	40	33	9	89	214	197	328	66	20.1	869	20	10.8	291	36	87.6	9	6	39
Montreal	41	30	11	93	242	240	374	72	19.2	1223	6	15	370	65	82.4	10	10	38
New Jersey	51	27	4	106	238	207	307	58	18.9	1038	20	12.9	324	65	79.9	12	3	44
Nashville	40	34	8	88	207	228	318	50	15.7	982	12	12.1	338	59	82.5	9	8	41
NY Islanders	26	47	9	61	198	274	320	54	16.9	1198	18	14.8	361	73	79.8	12	5	37
NY Rangers	43	30	9	95	200	212	346	48	13.9	1175	24	14.6	329	40	87.8	9	13	42
Ottawa	36	35	11	83	213	231	339	66	19.5	1084	14	13.4	346	64	81.5	8	5	46
Philadelphia	44	27	11	99	260	232	316	71	22.5	1408	26	17.5	393	67	83	16	1	43
Phoenix	36	39	7	79	205	249	344	50	14.5	1074	18	13.3	293	68	76.8	5	4	36
Pittsburgh	45	28	9	99	258	233	360	62	17.2	1106	8	13.6	347	60	82.7	7	11	46
San Jose	53	18	11	117	251	199	360	87	24.2	1037	16	12.8	306	51	83.3	12	10	46
St. Louis	41	31	10	92	227	227	351	72	20.5	1226	22	15.2	357	58	83.8	10	8	35
Tampa Bay	24	40	18	66	207	269	343	61	17.8	1280	26	15.9	405	89	78	4	8	34
Toronto	34	35	13	81	244	286	330	62	18.8	1113	12	13.7	308	78	74.7	6	7	40
Vancouver	45	27	10	100	243	213	357	67	18.8	1323	28	16.5	371	69	81.4	7	5	47
Washington	50	24	8	108	268	240	337	85	25.2	1021	20	12.7	387	75	80.6	7	9	45

W	Wins	PEN	Total penalty minutes including bench minutes
L	Losses during regular time	BMI	Total bench minor minutes
OTL	Overtime losses	AVG	Average penalty minutes per game
PTS	Points. Two points for winning a game, one point for a tie or losing in overtime, zero points for losing in regular time.	SHT	Total times short-handed. Measures opponent opportunities.
		PPGA	Power-play goals against
GF	Goals for	PKPCT	Penalty killing percentage. Measures a team's ability to prevent goals while its opponent is on a power play. Opponent opportunities minus power-play goals divided by opponent's opportunities.
GA	Goals against		
ADV	Total advantages. Power-play opportunities.		
PPGF	Power-play goals for. Goals scored while on power play.	SHGF	Short-handed goals for
PCTG	Power-play percentage. Power-play goals divided by total advantages.	SHGA	Short-handed goals against
		FG	Games scored first

Exercises for Section 12.2

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

12.2.1 Recall the regression of percent of body fat on height and waist from Exercise 12.1.1. The simple regression model of percent of body fat on height alone shows the following:

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  25.58078  14.15400   1.807  0.0719
Height       −0.09316   0.20119 −0.463  0.6438
```

a. Test whether the coefficient of height is statistically significant.
b. Looking at the model with both waist and height, test whether the coefficient of height is significant in this model.
c. Explain the discrepancy in your two answers.

12.2.2 Consider the linear regression model from Exercise 12.1.2. Is the second-order term necessary in the regression model?

12.2.3 Consider the regression model in Exercise 12.1.3 attempting to predict the percent of engineers in the workforce from various spending variables.
a. Are any of the variables useful for prediction? (Test an appropriate hypothesis.)
b. What percent of the variation in the percent of engineers is accounted for by the model?
c. What might you do next to create a better model?

12.2.4 You have fit a regression model with two regressors to a data set that has 20 observations. The total sum of squares is 1000 and the model sum of squares is 750.
a. What is the value of R^2 for this model?
b. What is the adjusted R^2 for this model?
c. What is the value of the F-statistic for testing the significance of regression? What conclusions would you draw about this model if $\alpha = 0.05$? What if $\alpha = 0.01$?
d. Suppose that you add a third regressor to the model and as a result, the model sum of squares is now 785. Does it seem to you that adding this factor has improved the model?

12.2.5 SS Consider the absorption index data in Exercise 12.1.4. The total sum of squares for y is $SS_T = 742.00$.
a. Test for significance of regression using $\alpha = 0.01$. What is the P-value for this test?
b. Test the hypothesis H_0: $\beta_1 = 0$ versus H_1: $\beta_1 \neq 0$ using $\alpha = 0.01$. What is the P-value for this test?
c. What conclusion can you draw about the usefulness of x_1 as a regressor in this model?

12.2.6 Consider the electric power consumption data in Exercise 12.1.6.
a. Test for significance of regression using $\alpha = 0.05$. What is the P-value for this test?
b. Use the t-test to assess the contribution of each regressor to the model. Using $\alpha = 0.05$, what conclusions can you draw?

12.2.7 Consider the regression model fit to the X-ray inspection data in Exercise 12.1.7. Use rads as the response.
a. Test for significance of regression using $\alpha = 0.05$. What is the P-value for this test?
b. Construct a t-test on each regression coefficient. What conclusions can you draw about the variables in this model? Use $\alpha = 0.05$.

12.2.8 Consider the wire bond pull strength data in Exercise 12.1.8.
a. Test for significance of regression using $\alpha = 0.05$. Find the P-value for this test. What conclusions can you draw?
b. Calculate the t-test statistic for each regression coefficient. Using $\alpha = 0.05$, what conclusions can you draw? Do all variables contribute to the model?

12.2.9 VS Consider the regression model fit to the gray range modulation data in Exercise 12.1.11. Use the useful range as the response.
a. Test for significance of regression using $\alpha = 0.05$. What is the P-value for this test?
b. Construct a t-test on each regression coefficient. What conclusions can you draw about the variables in this model? Use $\alpha = 0.05$.

12.2.10 Consider the regression model fit to the nisin extraction data in Exercise 12.1.10. Use nisin extraction as the response.
a. Test for significance of regression using $\alpha = 0.05$. What is the P-value for this test?
b. Construct a t-test on each regression coefficient. What conclusions can you draw about the variables in this model? Use $\alpha = 0.05$.
c. Comment on the effect of a small sample size to the tests in the previous parts.

12.2.11 SS Consider the bearing wear data in Exercise 12.1.13.
a. For the model with no interaction, test for significance of regression using $\alpha = 0.05$. What is the P-value for this test? What are your conclusions?
b. For the model with no interaction, compute the t-statistics for each regression coefficient. Using $\alpha = 0.05$, what conclusions can you draw?
c. For the model with no interaction, use the extra sum of squares method to investigate the usefulness of adding x_2 = load to a model that already contains x_1 = oil viscosity. Use $\alpha = 0.05$.
d. Refit the model with an interaction term. Test for significance of regression using $\alpha = 0.05$.
e. Use the extra sum of squares method to determine whether the interaction term contributes significantly to the model. Use $\alpha = 0.05$.
f. Estimate σ^2 for the interaction model. Compare this to the estimate of σ^2 from the model in part (a).

12.2.12 Data on National Hockey League team performance were presented in Exercise 12.1.12.
a. Test the model from this exercise for significance of regression using $\alpha = 0.05$. What conclusions can you draw?

b. Use the t-test to evaluate the contribution of each regressor to the model. Does it seem that all regressors are necessary? Use $\alpha = 0.05$.

c. Fit a regression model relating the number of games won to the number of goals for and the number of power-play goals for. Does this seem to be a logical choice of regressors, considering your answer to part (b)? Test this new model for significance of regression and evaluate the contribution of each regressor to the model using the t-test. Use $\alpha = 0.05$.

12.2.13 Data from a hospital patient satisfaction survey were presented in Exercise 12.1.5.

a. Test the model from this exercise for significance of regression. What conclusions can you draw if $\alpha = 0.05$? What if $\alpha = 0.01$?

b. Test the contribution of the individual regressors using the t-test. Does it seem that all regressors used in the model are really necessary?

Exercises for Sections 12.3 and 12.4

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

12.4.1 Using the regression model from Exercise 12.1.1,

a. Find a 95% confidence interval for the coefficient of height.

b. Find a 95% confidence interval for the mean percent of body fat for a man with a height of 72 in. and waist of 34 in.

c. Find a 95% prediction interval for the percent of body fat for a man with the same height and waist as in part (b).

d. Which interval is wider, the confidence interval or the prediction interval? Explain briefly.

e. Given your answer to part (c), do you believe that this is a useful model for predicting body fat? Explain briefly.

12.4.2 Using the second-order polynomial regression model from Exercise 12.1.2,

a. Find a 95% confidence interval on both the first-order and the second-order term in this model.

b. Is zero in the confidence interval you found for the second-order term in part (a)? What does that fact tell you about the contribution of the second-order term to the model?

c. Refit the model with only the first-order term. Find a 95% confidence interval on this term. Is this interval longer or shorter than the confidence interval that you found on this term in part (a)?

12.4.3 Consider the soil absorption data in Exercise 12.1.4.

a. Find 95% confidence intervals on the regression coefficients.

b. Find a 95% confidence interval on the mean soil absorption index when $x_1 = 200$ and $x_2 = 50$.

c. Find a 95% prediction interval on the soil absorption index when $x_1 = 200$ and $x_2 = 50$.

12.4.4 Consider the electric power consumption data in Exercise 12.1.6.

a. Find 95% confidence intervals on β_1, β_2, β_3, and β_4.

b. Find a 95% confidence interval on the mean of Y when $x_1 = 75$, $x_2 = 24$, $x_3 = 90$, and $x_4 = 98$.

c. Find a 95% prediction interval on the power consumption when $x_1 = 75$, $x_2 = 24$, $x_3 = 90$, and $x_4 = 98$.

12.4.5 Consider the regression model fit to the X-ray inspection data in Exercise 12.1.7. Use rads as the response.

a. Calculate 95% confidence intervals on each regression coefficient.

b. Calculate a 99% confidence interval on mean rads at 15 milliamps and 1 second on exposure time.

c. Calculate a 99% prediction interval on rads for the same values of the regressors used in part (b).

12.4.6 WP SS Consider the wire bond pull strength data in Exercise 12.1.8.

a. Find a 95% confidence interval on the regression coefficients.

b. Find a 95% confidence interval on mean pull strength when $x_2 = 20$, $x_3 = 30$, $x_4 = 90$, and $x_5 = 2.0$.

c. Find a 95% prediction interval on pull strength when $x_2 = 20$, $x_3 = 30$, $x_4 = 90$, and $x_5 = 2.0$.

12.4.7 Consider the regression model fit to the gray range modulation data in Exercise 12.1.11. Use the useful range as the response.

a. Calculate 99% confidence intervals on each regression coefficient.

b. Calculate a 99% confidence interval on mean useful range when brightness = 70 and contrast = 80.

c. Calculate a prediction interval on useful range for the same values of the regressors used in part (b).

d. Calculate a 99% confidence interval and a 99% prediction interval on useful range when brightness = 50 and contrast = 25. Compare the widths of these intervals to those calculated in parts (b) and (c). Explain any differences in widths.

12.4.8 WP SS Consider the regression model fit to the nisin extraction data in Exercise 12.1.10.

a. Calculate 95% confidence intervals on each regression coefficient.

b. Calculate a 95% confidence interval on mean nisin extraction when $x_1 = 15.5$ and $x_2 = 16$.

c. Calculate a prediction interval on nisin extraction for the same values of the regressors used in part (b).

d. Comment on the effect of a small sample size to the widths of these intervals.

12.4.9 Consider the NHL data in Exercise 12.1.12.

a. Find a 95% confidence interval on the regression coefficient for the variable GF.

b. Fit a simple linear regression model relating the response variable to the regressor *GF*.
c. Find a 95% confidence interval on the slope for the simple linear regression model from part (b).
d. Compare the lengths of the two confidence intervals computed in parts (a) and (c). Which interval is shorter? Does this tell you anything about which model is preferable?

Exercises for Section 12.5

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

12.5.1 WP Consider the regression model fit to the X-ray inspection data in Exercise 12.1.7. Use rads as the response.
a. What proportion of total variability is explained by this model?
b. Construct a normal probability plot of the residuals. What conclusion can you draw from this plot?
c. Plot the residuals versus $\hat{y}$ and versus each regressor, and comment on model adequacy.
d. Calculate Cook's distance for the observations in this data set. Are there any influential points in these data?

12.5.2 Consider the electric power consumption data in Exercise 12.1.6.
a. Calculate R^2 for this model. Interpret this quantity.
b. Plot the residuals versus $\hat{y}$ and versus each regressor. Interpret this plot.
c. Construct a normal probability plot of the residuals and comment on the normality assumption.

12.5.3 WP SS VS Consider the regression model fit to the coal and limestone mixture data in Exercise 12.1.9. Use density as the response.
a. What proportion of total variability is explained by this model?
b. Construct a normal probability plot of the residuals. What conclusion can you draw from this plot?
c. Plot the residuals versus $\hat{y}$ and versus each regressor, and comment on model adequacy.
d. Calculate Cook's distance for the observations in this data set. Are there any influential points in these data?

12.5.4 Consider the regression model fit to the nisin extraction data in Exercise 12.1.10.
a. What proportion of total variability is explained by this model?
b. Construct a normal probability plot of the residuals. What conclusion can you draw from this plot?
c. Plot the residuals versus $\hat{y}$ and versus each regressor, and comment on model adequacy.
d. Calculate Cook's distance for the observations in this data set. Are there any influential points in these data?

12.5.5 Consider the regression model fit to the gray range modulation data in Exercise 12.1.11. Use the useful range as the response.
a. What proportion of total variability is explained by this model?
b. Construct a normal probability plot of the residuals. What conclusion can you draw from this plot?
c. Plot the residuals versus $\hat{y}$ and versus each regressor, and comment on model adequacy.
d. Calculate Cook's distance for the observations in this data set. Are there any influential points in these data?

12.5.6 Consider the regression model for the NHL data from Exercise 12.1.12.
a. Fit a model using *GF* as the only regressor.
b. How much variability is explained by this model?
c. Plot the residuals versus $\hat{y}$ and comment on model adequacy.
d. Plot the residuals from part (a) versus *PPGF*, the points scored while in power play. Does this indicate that the model would be better if this variable were included?

12.5.7 Consider the bearing wear data in Exercise 12.1.13.
a. Find the value of R^2 when the model uses the regressors x_1 and x_2.
b. What happens to the value of R^2 when an interaction term x_1x_2 is added to the model? Does this necessarily imply that adding the interaction term is a good idea?

Exercises for Section 12.6

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

12.6.1 An article titled "A Method for Improving the Accuracy of Polynomial Regression Analysis" in the *Journal of Quality Technology* (1971, Vol. 3(4), pp. 149–155) reported the following data on y = ultimate shear strength of a rubber compound (psi) and x = cure temperature (°F).

y	770	800	840	810
x	280	284	292	295
y	735	640	590	560
x	298	305	308	315

a. Fit a second-order polynomial to these data.
b. Test for significance of regression using $\alpha = 0.05$.
c. Test the hypothesis that $\beta_{11} = 0$ using $\alpha = 0.05$.
d. Compute the residuals from part (a) and use them to evaluate model adequacy.

12.6.2 Consider the following data, which result from an experiment to determine the effect of x = test time in hours at a particular temperature on y = change in oil viscosity:

y	−1.42	−1.39	−1.55	−1.89	−2.43
x	.25	.50	.75	1.00	1.25
y	−3.15	−4.05	−5.15	−6.43	−7.89
x	1.50	1.75	2.00	2.25	2.50

a. Fit a second-order polynomial to the data.
b. Test for significance of regression using $\alpha = 0.05$.
c. Test the hypothesis that $\beta_{11} = 0$ using $\alpha = 0.05$.
d. Compute the residuals from part (a) and use them to evaluate model adequacy.

12.6.3 SS The following data were collected during an experiment to determine the change in thrust efficiency (y, in percent) as the divergence angle of a rocket nozzle (x) changes:

y	24.60	24.71	23.90	39.50	39.60	57.12
x	4.0	4.0	4.0	5.0	5.0	6.0
y	67.11	67.24	67.15	77.87	80.11	84.67
x	6.5	6.5	6.75	7.0	7.1	7.3

a. Test for significance of regression using $\alpha = 0.05$.
b. Test for significance of regression and lack of fit using $\alpha = 0.05$.
c. Test the hypothesis that $\beta_{11} = 0$, using $\alpha = 0.05$.
d. Plot the residuals and comment on model adequacy.
e. Fit a cubic model, and test for the significance of the cubic term using $\alpha = 0.05$.

12.6.4 An article in the *Journal of Pharmaceuticals Sciences* ["Statistical Analysis of the Extended Hansen Method Using the Bootstrap Technique" (1991, Vol. 80, pp. 971–977)] presents data on the observed mole fraction solubility of a solute at a constant temperature and the dispersion, dipolar, and hydrogen-bonding Hansen partial solubility parameters. The data are shown in the following table, where y is the negative logarithm of the mole fraction solubility, x_1 is the dispersion partial solubility, x_2 is the dipolar partial solubility, and x_3 is the hydrogen-bonding partial solubility.

Observation Number	y	x_1	x_2	x_3
1	0.22200	7.3	0.0	0.0
2	0.39500	8.7	0.0	0.3
3	0.42200	8.8	0.7	1.0
4	0.43700	8.1	4.0	0.2
5	0.42800	9.0	0.5	1.0
6	0.46700	8.7	1.5	2.8
7	0.44400	9.3	2.1	1.0
8	0.37800	7.6	5.1	3.4
9	0.49400	10.0	0.0	0.3
10	0.45600	8.4	3.7	4.1
11	0.45200	9.3	3.6	2.0
12	0.11200	7.7	2.8	7.1
13	0.43200	9.8	4.2	2.0
14	0.10100	7.3	2.5	6.8
15	0.23200	8.5	2.0	6.6
16	0.30600	9.5	2.5	5.0
17	0.09230	7.4	2.8	7.8
18	0.11600	7.8	2.8	7.7
19	0.07640	7.7	3.0	8.0
20	0.43900	10.3	1.7	4.2
21	0.09440	7.8	3.3	8.5
22	0.11700	7.1	3.9	6.6
23	0.07260	7.7	4.3	9.5
24	0.04120	7.4	6.0	10.9
25	0.25100	7.3	2.0	5.2
26	0.00002	7.6	7.8	20.7

a. Fit the model $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \varepsilon$.
b. Test for significance of regression using $\alpha = 0.05$.
c. Plot the residuals and comment on model adequacy.
d. Use the extra sum of squares method to test the contribution of the second-order terms using $\alpha = 0.05$.

12.6.5 Consider the electric power data in Exercise 12.1.6. Build regression models for the data using the following techniques:
a. All possible regressions. Find the minimum C_p and minimum MS_E equations.
b. Stepwise regression.
c. Forward selection.
d. Backward elimination.
e. Comment on the models obtained. Which model would you prefer?

12.6.6 Consider the X-ray inspection data in Exercise 12.1.7. Use rads as the response. Build regression models for the data using the following techniques:
a. All possible regressions.
b. Stepwise regression.
c. Forward selection.
d. Backward elimination.
e. Comment on the models obtained. Which model would you prefer? Why?

12.6.7 Consider the wire bond pull strength data in Exercise 12.1.8. Build regression models for the data using the following methods:

a. All possible regressions. Find the minimum C_p and minimum MS_E equations.
b. Stepwise regression.
c. Forward selection.
d. Backward elimination.
e. Comment on the models obtained. Which model would you prefer?

12.6.8 Consider the regression model fit to the coal and limestone mixture data in Exercise 12.1.9. Use density as the response. Build regression models for the data using the following techniques:

a. All possible regressions.
b. Stepwise regression.
c. Forward selection.
d. Backward elimination.
e. Comment on the models obtained. Which model would you prefer? Why?

12.6.9 Consider the nisin extraction data in Exercise 12.1.10. Build regression models for the data using the following techniques:

a. All possible regressions.
b. Stepwise regression.
c. Forward selection.
d. Backward elimination.
e. Comment on the models obtained. Which model would you prefer? Why?

12.6.10 Consider the gray range modulation data in Exercise 12.1.11. Use the useful range as the response. Build regression models for the data using the following techniques:

a. All possible regressions.
b. Stepwise regression.
c. Forward selection.
d. Backward elimination.
e. Comment on the models obtained. Which model would you prefer? Why?

12.6.11 Consider the NHL data in Exercise 12.1.12. Build regression models for these data with regressors *GF* through *FG* using the following methods:

a. All possible regressions. Find the minimum *Cp* and minimum MS_E equations.
b. Stepwise regression.
c. Forward selection.
d. Backward elimination.
e. Which model would you prefer?

12.6.12 WP SS When fitting polynomial regression models, we often subtract $\bar{x}$ from each $\bar{x}$ value to produce a "centered" regressor $x' = x - \bar{x}$. This reduces the effects of dependencies among the model terms and often leads to more accurate estimates of the regression coefficients. Using the data from Exercise 12.6.1, fit the model $Y = \beta_0^* + \beta_1^* x' + \beta_{11}^* (x')^2 + \varepsilon$.

a. Use the results to estimate the coefficients in the uncentered model $Y = \beta_0 + \beta_1 x + \beta_{11} x^2 + \varepsilon$. Predict y when $x = 285°\text{F}$. Suppose that you use a standardized variable $x' = (x - \bar{x})/s_x$ where s_x is the standard deviation of x in constructing a polynomial regression model. Fit the model $Y = \beta_0^* + \beta_1^* x' + \beta_{11}^* (x')^2 + \varepsilon$.
b. What value of y do you predict when $x = 285°\text{F}$?
c. Estimate the regression coefficients in the unstandardized model $Y = \beta_0 + \beta_1 x + \beta_{11} x^2 + \varepsilon$.
d. What can you say about the relationship between SS_E and R^2 for the standardized and unstandardized models?
e. Suppose that $y' = (y - \bar{y})/s_y$ is used in the model along with x'. Fit the model and comment on the relationship between SS_E and R^2 in the standardized model and the unstandardized model.

12.6.13 WP Consider the data in Exercise 12.6.4. Use all the terms in the full quadratic model as the candidate regressors.

a. Use forward selection to identify a model.
b. Use backward elimination to identify a model.
c. Compare the two models obtained in parts (a) and (b). Which model would you prefer and why?

12.6.14 We have used a sample of 30 observations to fit a regression model. The full model has nine regressors, the variance estimate is $\hat{\sigma}^2 = MS_E = 100$ and $R^2 = 0.92$.

a. Calculate the F-statistic for testing significance of regression. Using $\alpha = 0.05$, what would you conclude?
b. Suppose that we fit another model using only four of the original regressors and that the error sum of squares for this new model is 2200. Find the estimate of σ^2 for this new reduced model. Would you conclude that the reduced model is superior to the old one? Why?
c. Find the value of C_p for the reduced model in part (b). Would you conclude that the reduced model is better than the old model?

12.6.15 A sample of 25 observations is used to fit a regression model in seven variables. The estimate of σ^2 for this full model is $MS_E = 10$.

a. A forward selection algorithm has put three of the original seven regressors in the model. The error sum of squares for the three-variable model is $SS_E = 300$. Based on C_p, would you conclude that the three-variable model has any remaining bias?
b. After looking at the forward selection model in part (a), suppose you could add one more regressor to the model. This regressor will reduce the error sum of squares to 275. Will the addition of this variable improve the model? Why?

Supplemental Exercises for Chapter 12

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

12.S8 WP Consider the following computer output.

```
The regression equation is

Y = 517 + 11.5 x1 − 8.14 x2 + 10.9 x3

Predictor   Coef      SE Coef   T       P
Constant    517.46    11.76     ?       ?
x1          11.4720   ?         36.50   ?
x2          −8.1378   0.1969    ?       ?
x3          10.8565   0.6652    ?       ?

S = 10.2560    R-Sq = ?    R-Sq (adj) = ?

Analysis of Variance

Source                DF  SS      MS      F  P
Regression            ?   347300  115767  ?  ?
Residual error        16  ?       105
Total                 19  348983
```

a. Fill in the missing values. Use bounds for the P-values.
b. Is the overall model significant at $\alpha = 0.05$? Is it significant at $\alpha = 0.01$?
c. Discuss the contribution of the individual regressors to the model.

12.S9 WP SS Consider the inverter data in Exercise 12.S10. Delete observation 2 from the original data. Define new variables as follows: $y^* = \ln y$, $x_1^* = 1/\sqrt{x_1}$, $x_2^* = \sqrt{x_2}$, $x_3^* = 1/\sqrt{x_3}$, and $x_4^* = \sqrt{x_4}$.

a. Fit a regression model using these transformed regressors (do not use x_5 or x_6).
b. Test the model for significance of regression using $\alpha = 0.05$. Use the t-test to investigate the contribution of each variable to the model ($\alpha = 0.05$). What are your conclusions?
c. Plot the residuals versus $\hat{y}^*$ and versus each of the transformed regressors. Comment on the plots.

12.S10 WP Transient points of an electronic inverter are influenced by many factors. Shown below is the data on the transient point (y, in volts) of PMOS-NMOS inverters and five candidate regressors: x_1 = width of the NMOS device, x_2 = length of the NMOS device, x_3 = width of the PMOS device, x_4 = length of the PMOS device, and x_5 = temperature (°C).

Observation Number	x_1	x_2	x_3	x_4	x_5	y
1	3	3	3	3	0	0.787
2	8	30	8	8	0	0.293
3	3	6	6	6	0	1.710
4	4	4	4	12	0	0.203
5	8	7	6	5	0	0.806
6	10	20	5	5	0	4.713
7	8	6	3	3	25	0.607
8	6	24	4	4	25	9.107
9	4	10	12	4	25	9.210
10	16	12	8	4	25	1.365
11	3	10	8	8	25	4.554
12	8	3	3	3	25	0.293
13	3	6	3	3	50	2.252
14	3	8	8	3	50	9.167
15	4	8	4	8	50	0.694
16	5	2	2	2	50	0.379
17	2	2	2	3	50	0.485
18	10	15	3	3	50	3.345
19	15	6	2	3	50	0.208
20	15	6	2	3	75	0.201
21	10	4	3	3	75	0.329
22	3	8	2	2	75	4.966
23	6	6	6	4	75	1.362
24	2	3	8	6	75	1.515
25	3	3	8	8	75	0.751

a. Fit a multiple linear regression model that uses all regressors to these data. Test for significance of regression using $\alpha = 0.01$. Find the P-value for this test and use it to draw your conclusions.
b. Test the contribution of each variable to the model using the t-test with $\alpha = 0.05$. What are your conclusions?
c. Delete x_5 from the model. Test the new model for significance of regression. Also test the relative contribution of each regressor to the new model with the t-test. Using $\alpha = 0.05$, what are your conclusions?
d. Notice that the MS_E for the model in part (c) is smaller than the MS_E for the full model in part (a). Explain why this has occurred.

e. Calculate the studentized residuals. Do any of these seem unusually large?

f. Suppose that you learn that the second observation was recorded incorrectly. Delete this observation and refit the model using x_1, x_2, x_3, and x_4 as the regressors. Notice that the R^2 for this model is considerably higher than the R^2 for either of the models fitted previously. Explain why the R^2 for this model has increased.

g. Test the model from part (f) for significance of regression using $\alpha = 0.05$. Also investigate the contribution of each regressor to the model using the t-test with $\alpha = 0.05$. What conclusions can you draw?

h. Plot the residuals from the model in part (f) versus $\hat{y}$ and versus each of the regressors x_1, x_2, x_3, and x_4. Comment on the plots.

12.S11 WP A multiple regression model was used to relate y = viscosity of a chemical product to x_1 = temperature and x_2 = reaction time. The data set consisted of $n = 15$ observations.

a. The estimated regression coefficients were $\hat{\beta}_0 = 300.00$, $\hat{\beta}_1 = 0.85$, and $\hat{\beta}_2 = 10.40$. Calculate an estimate of mean viscosity when $x_1 = 100°\text{F}$ and $x_2 = 2$ hours.

b. The sums of squares were $SS_T = 1230.50$ and $SS_E = 120.30$. Test for significance of regression using $\alpha = 0.05$. What conclusion can you draw?

c. What proportion of total variability in viscosity is accounted for by the variables in this model?

d. Suppose that another regressor, x_3 = stirring rate, is added to the model. The new value of the error sum of squares is $SS_E = 117.20$. Has adding the new variable resulted in a smaller value of MS_E? Discuss the significance of this result.

e. Calculate an F-statistic to assess the contribution of x_3 to the model. Using $\alpha = 0.05$, what conclusions do you reach?

12.S12 Consider the electronic inverter data in Exercises 12.S9 and 12.S10. Define the response and regressor variables as in Exercise 12.S9, and delete the second observation in the sample.

a. Use all possible regressions to find the equation that minimizes C_p.

b. Use all possible regressions to find the equation that minimizes MS_E.

c. Use stepwise regression to select a subset regression model.

d. Compare the models you have obtained.

12.S13 WP An article in the *Journal of the American Ceramics Society* ["In Situ Vitrification and the Effects of Soil Additives—A Mixture Experiment Case Study" (1992, Vol. 75, pp. 112–116)] described a process for immobilizing chemical or nuclear wastes in soil by dissolving the contaminated soil into a glass block. The authors mix CaO and Na_2O with soil and model viscosity and electrical conductivity. The electrical conductivity model involves six regressors, and the sample consists of $n = 14$ observations.

a. For the six-regressor model, suppose that $SS_T = 0.50$ and $R^2 = 0.94$. Find SS_E and SS_R, and use this information to test for significance of regression with $\alpha = 0.05$. What are your conclusions?

b. Suppose that one of the original regressors is deleted from the model, resulting in $R^2 = 0.92$. What can you conclude about the contribution of the variable that was removed? Answer this question by calculating an F-statistic.

c. Does deletion of the regressor variable in part (b) result in a smaller value of MS_E for the five-variable model, in comparison to the original six-variable model? Comment on the significance of your answer.

12.S14 Exercise 12.1.5 introduced the hospital patient satisfaction survey data. One of the variables in that data set is a categorical variable indicating whether the patient is a medical patient or a surgical patient. Fit a model including this indicator variable to the data using all three of the other regressors. Is there any evidence that the service the patient is on (medical versus surgical) has an impact on the reported satisfaction?

12.S15 Consider the following inverse model matrix.

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.125 & 0 & 0 & 0 \\ 0 & 0.125 & 0 & 0 \\ 0 & 0 & 0.125 & 0 \\ 0 & 0 & 0 & 0.125 \end{bmatrix}$$

a. How many regressors are in this model?

b. What was the sample size?

c. Notice the special diagonal structure of the matrix. What does that tell you about the columns in the original $\mathbf{X}$ matrix?

Chapter 13 Exercises

Exercises for Section 13.2

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

13.2.1 VS Consider the following computer output.

Source	DF	SS	MS	F	*P*-value
Factor	?	117.4	39.1	?	?
Error	16	396.8	?		
Total	19	514.2			

a. How many levels of the factor were used in this experiment?
b. How many replicates did the experimenter use?
c. Fill in the missing information in the ANOVA table. Use bounds for the *P*-value.
d. What conclusions can you draw about differences in the factor-level means?

13.2.2 Consider the following computer output for an experiment. The factor was tested over four levels.

Source	DF	SS	MS	F	*P*-value
Factor	?	?	330.4716	4.42	?
Error	?	?	?		
Total	31	?			

a. How many replicates did the experimenter use?
b. Fill in the missing information in the ANOVA table. Use bounds for the *P*-value.
c. What conclusions can you draw about differences in the factor-level means?

13.2.3 WP In *Design and Analysis of Experiments*, 9th edition (Wiley, 2017), D. C. Montgomery described an experiment in which the tensile strength of a synthetic fiber was of interest to the manufacturer. It is suspected that strength is related to the percentage of cotton in the fiber. Five levels of cotton percentage were used and five replicates were run in random order, resulting in the following data.

Cotton Percentage	Tensile Strength				
15	7	7	15	11	9
20	12	17	12	18	18
25	14	18	18	19	19
30	19	25	22	19	23
35	7	10	11	15	11

a. Does cotton percentage affect breaking strength? Draw comparative box plots and perform an analysis of variance. Use $\alpha = 0.05$.
b. Plot average tensile strength against cotton percentage and interpret the results.
c. Analyze the residuals and comment on model adequacy.

13.2.4 An article in *Nature* describes an experiment to investigate the effect on consuming chocolate on cardiovascular health ("Plasma Antioxidants from Chocolate," 2003, Vol. 424, p. 1013). The experiment consisted of using three different types of chocolates: 100 g of dark chocolate, 100 g of dark chocolate with 200 ml of full-fat milk, and 200 g of milk chocolate. Twelve subjects were used, seven women and five men with an average age range of 32.2 ± 1 years, an average weight of 65.8 ± 3.1 kg, and body mass index of 21.9 ± 0.4 kg m^{-2}. On different days, a subject consumed one of the chocolate-factor levels, and one hour later total antioxidant capacity of that person's blood plasma was measured in an assay. Data similar to those summarized in the article follow.

DC		DC + MK		MC	
118.8	115.8	105.4	100.0	102.1	102.8
122.6	115.1	101.1	99.8	105.8	98.7
115.6	116.9	102.7	102.6	99.6	94.7
113.6	115.4	97.1	100.9	102.7	97.8
119.5	115.6	101.9	104.5	98.8	99.7
115.9	107.9	98.9	93.5	100.9	98.6

a. Construct comparative box plots and study the data. What visual impression do you have from examining these plots?
b. Analyze the experimental data using an ANOVA. If $\alpha = 0.05$, what conclusions would you draw? What would you conclude if $\alpha = 0.01$?
c. Is there evidence that the dark chocolate increases the mean antioxidant capacity of the subjects' blood plasma?
d. Analyze the residuals from this experiment.

13.2.5 WP In "Orthogonal Design for Process Optimization and Its Application to Plasma Etching" (*Solid State Technology*, 1987), G. Z. Yin and D. W. Jillie described an experiment to determine the effect of C_2F_6 flow rate on the uniformity of the etch on a silicon wafer used in integrated circuit manufacturing. Three flow rates are used in the experiment, and the resulting uniformity (in percent) for six replicates follows.

C_2F_6 Flow (SCCM)	Uniformity (%)					
125	2.7	4.6	2.6	3.0	3.2	3.8
160	4.9	4.6	5.0	4.2	3.6	4.2
200	4.6	3.4	2.9	3.5	4.1	5.1

a. Does C_2F_6 flow rate affect etch uniformity? Construct box plots to compare the factor levels and perform the analysis of variance. Use $\alpha = 0.05$.
b. Do the residuals indicate any problems with the underlying assumptions?

13.2.6 An article in *Environment International* ["Influence of Water Temperature and Shower Head Orifice Size on the Release of Radon During Showering" (1992, Vol. 18(4), pp. 363–369)] described an experiment in which the amount of radon released in showers was investigated. Radon-enriched water was used in the experiment, and six different orifice diameters were tested in shower heads. The data from the experiment are shown in the following table.

Orifice Diameter	Radon Released (%)			
0.37	80	83	83	85
0.51	75	75	79	79
0.71	74	73	76	77
1.02	67	72	74	74
1.40	62	62	67	69
1.99	60	61	64	66

a. Does the size of the orifice affect the mean percentage of radon released? Use $\alpha = 0.05$.
b. Find the P-value for the F-statistic in part (a).
c. Analyze the residuals from this experiment.
d. Find a 95% confidence interval on the mean percent of radon released when the orifice diameter is 1.40.

13.2.7 WP SS An article in the *ACI Materials Journal* (1987, Vol. 84, pp. 213–216) described several experiments investigating the rodding of concrete to remove entrapped air. A 3-inch × 6-inch cylinder was used, and the number of times this rod was used is the design variable. The resulting compressive strength of the concrete specimen is the response. The data are shown in the following table.

Rodding Level	Compressive Strength		
10	1530	1530	1440
15	1610	1650	1500
20	1560	1730	1530
25	1500	1490	1510

a. Is there any difference in compressive strength due to the rodding level?
b. Find the P-value for the F-statistic in part (a).
c. Analyze the residuals from this experiment. What conclusions can you draw about the underlying model assumptions?

13.2.8 An article in the *Materials Research Bulletin* [1991, Vol. 26(11), pp. 1151–1165] investigated four different methods of preparing the superconducting compound $PbMo_6S_8$. The authors contend that the presence of oxygen during the preparation process affects the material's superconducting transition temperature T_c. Preparation methods 1 and 2 use techniques that are designed to eliminate the presence of oxygen, and methods 3 and 4 allow oxygen to be present. Five observations on T_c (in °K) were made for each method, and the results are as follows:

Preparation Method	Transition Temperature T_c(°K)				
1	14.8	14.8	14.7	14.8	14.9
2	14.6	15.0	14.9	14.8	14.7
3	12.7	11.6	12.4	12.7	12.1
4	14.2	14.4	14.4	12.2	11.7

a. Is there evidence to support the claim that the presence of oxygen during preparation affects the mean transition temperature? Use $\alpha = 0.05$.
b. What is the P-value for the F-test in part (a)?
c. Analyze the residuals from this experiment.
d. Find a 95% confidence interval on mean T_c when method 1 is used to prepare the material.

13.2.9 VS A paper in the *Journal of the Association of Asphalt Paving Technologists* (1990, Vol. 59) described an experiment to determine the effect of air voids on percentage retained strength of asphalt. For purposes of the experiment, air voids are controlled at three levels; low (2–4%), medium (4–6%), and high (6–8%). The data are shown in the following table.

Air Voids	Retained Strength (%)							
Low	106	90	103	90	79	88	92	95
Medium	80	69	94	91	70	83	87	83
High	78	80	62	69	76	85	69	85

a. Do the different levels of air voids significantly affect mean retained strength? Use $\alpha = 0.01$.
b. Find the P-value for the F-statistic in part (a).
c. Analyze the residuals from this experiment.
d. Find a 95% confidence interval on mean retained strength where there is a high level of air voids.
e. Find a 95% confidence interval on the difference in mean retained strength at the low and high levels of air voids.

13.2.10 An article in *Quality Engineering* ["Estimating Sources of Variation: A Case Study from Polyurethane Product Research" (1999–2000, Vol. 12, pp. 89–96)] reported a study on the effects of additives on final polymer properties. In this case, polyurethane additives were referred to as cross-linkers. The average domain spacing was the measurement of the polymer property. The data are as follows:

Cross-Linker Level	Domain Spacing (nm)					
−1	8.2	8	8.2	7.9	8.1	8
−0.75	8.3	8.4	8.3	8.2	8.3	8.1
−0.5	8.9	8.7	8.9	8.4	8.3	8.5
0	8.5	8.7	8.7	8.7	8.8	8.8
0.5	8.8	9.1	9.0	8.7	8.9	8.5
1	8.6	8.5	8.6	8.7	8.8	8.8

a. Is there a difference in the cross-linker level? Draw comparative box plots and perform an analysis of variance. Use $\alpha = 0.05$.
b. Find the P-value of the test. Estimate the variability due to random error.
c. Plot average domain spacing against cross-linker level and interpret the results.
d. Analyze the residuals from this experiment and comment on model adequacy.

13.2.11 WP SS An article in *Journal of Food Science* ["Prevention of Potato Spoilage During Storage by Chlorine Dioxide" (2001, Vol. 66(3), pp. 472–477)] reported on a study of potato spoilage based on different conditions of acidified oxine (AO), which is a mixture of chlorite and chlorine dioxide. The data follow:

AO Solution (ppm)	% Spoilage		
50	100	50	60
100	60	30	30
200	60	50	29
400	25	30	15

a. Do the AO solutions differ in the spoilage percentage? Use $\alpha = 0.05$.
b. Find the P-value of the test. Estimate the variability due to random error.

c. Plot average spoilage against AO solution and interpret the results. Which AO solution would you recommend for use in practice?

d. Analyze the residuals from this experiment.

13.2.12 An article in *Scientia Iranica* ["Tuning the Parameters of an Artificial Neural Network (ANN) Using Central Composite Design and Genetic Algorithm" (2011, Vol. 18(6), pp. 1600–1608)] described a series of experiments to tune parameters in artificial neural networks. One experiment considered the relationship between model fitness [measured by the square root of mean square error (RMSE) on a separate test set of data] and model complexity that were controlled by the number of nodes in the two hidden layers. The following data table (extracted from a much larger data set) contains three different ANNs: ANN1 has 33 nodes in layer 1 and 30 nodes in layer 2, ANN2 has 49 nodes in layer 1 and 45 nodes in layer 2, and ANN3 has 17 nodes in layer 1 and 15 nodes in layer 2.

ANN1	ANN2	ANN3
0.0121	0.0031	0.1562
0.0132	0.0006	0.2227
0.0011	0	0.0953
0.0023	0	0.8911
0.0391	0.022	1.3892
0.0054	0.0019	0.0154
0.0003	0.0007	1.7916
0.0014	0	0.1992

a. Construct a box plot to compare the different ANNs.

b. Perform the analysis of variance with $\alpha = 0.05$. What is the *P*-value?

c. Analyze the residuals from the experiment.

d. Calculate a 95% confidence interval on RMSE for ANN2.

For each of the following exercises, use the previous data to complete these parts.

a. Apply Fisher's LSD method with $\alpha = 0.05$ and determine which levels of the factor differ.

b. Use the graphical method to compare means described in this section and compare your conclusions to those from Fisher's LSD method.

13.2.13 Cotton percentage in Exercise 13.2.3. Use $\alpha = 0.05$.

13.2.14 Chocolate type in Exercise 13.2.4. Use $\alpha = 0.05$.

13.2.15 WP Flow rate in Exercise 13.2.5. Use $\alpha = 0.01$.

13.2.16 Preparation method in Exercise 13.2.8. Use $\alpha = 0.05$.

13.2.17 Air voids in Exercise 13.2.9. Use $\alpha = 0.05$.

13.2.18 Cross-linker Exercise 13.2.10. Use $\alpha = 0.05$.

Exercises for Section 13.3

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

13.3.1 WP An article in the *Journal of the Electrochemical Society* [1992, Vol. 139(2), pp. 524–532)] describes an experiment to investigate the low-pressure vapor deposition of polysilicon. The experiment was carried out in a large-capacity reactor at Sematech in Austin, Texas. The reactor has several wafer positions, and four of these positions were selected at random. The response variable is film thickness uniformity. Three replicates of the experiment were run, and the data are as follows:

a. Is there a difference in the wafer positions? Use $\alpha = 0.05$.

b. Estimate the variability due to wafer positions.

c. Estimate the random error component.

d. Analyze the residuals from this experiment and comment on model adequacy.

Wafer Position	Uniformity		
1	2.76	5.67	4.49
2	1.43	1.70	2.19
3	2.34	1.97	1.47
4	0.94	1.36	1.65

13.3.2 WP VS A textile mill has a large number of looms. Each loom is supposed to provide the same output of cloth per minute. To investigate this assumption, five looms are chosen at random, and their output is measured at different times. The following data are obtained:

Loom	Output (lb/min)				
1	4.0	4.1	4.2	4.0	4.1
2	3.9	3.8	3.9	4.0	4.0
3	4.1	4.2	4.1	4.0	3.9
4	3.6	3.8	4.0	3.9	3.7
5	3.8	3.6	3.9	3.8	4.0

a. Are the looms similar in output? Use $\alpha = 0.05$.

b. Estimate the variability between looms.

c. Estimate the experimental error variance.

d. Analyze the residuals from this experiment and check for model adequacy.

13.3.3 In the book *Bayesian Inference in Statistical Analysis* (1973, John Wiley and Sons) by Box and Tiao, the total product yield for five samples was determined randomly selected from each of six randomly chosen batches of raw material.

Batch	Yield (in grams)				
1	1545	1440	1440	1520	1580
2	1540	1555	1490	1560	1495
3	1595	1550	1605	1510	1560
4	1445	1440	1595	1465	1545
5	1595	1630	1515	1635	1625
6	1520	1455	1450	1480	1445

a. Do the different batches of raw material significantly affect mean yield? Use $\alpha = 0.01$.
b. Estimate the variability between batches.
c. Estimate the variability between samples within batches.
d. Analyze the residuals from this experiment and check for model adequacy.

13.3.4 WP An article in the *Journal of Quality Technology* [1981, Vol. 13(2), pp. 111–114)] described an experiment that investigated the effects of four bleaching chemicals on pulp brightness. These four chemicals were selected at random from a large population of potential bleaching agents. The data are as follows:

Chemical	Pulp Brightness				
1	77.199	74.466	92.746	76.208	82.876
2	80.522	79.306	81.914	80.346	73.385
3	79.417	78.017	91.596	80.802	80.626
4	78.001	78.358	77.544	77.364	77.386

a. Is there a difference in the chemical types? Use $\alpha = 0.05$.
b. Estimate the variability due to chemical types.
c. Estimate the variability due to random error.
d. Analyze the residuals from this experiment and comment on model adequacy.

13.3.5 WP SS Consider the vapor-deposition experiment described in Exercise 13.3.1.
a. Estimate the total variability in the uniformity response.
b. How much of the total variability in the uniformity response is due to the difference between positions in the reactor?
c. To what level could the variability in the uniformity response be reduced if the position-to-position variability in the reactor could be eliminated? Do you believe this is a substantial reduction?

13.3.6 Consider the cloth experiment described in Exercise 13.3.2.
a. Estimate the total variability in the output response.
b. 13.67 How much of the total variability in the output response is due to the difference between looms?
c. To what level could the variability in the output response be reduced if the loom-to-loom variability could be eliminated? Do you believe this is a significant reduction?

Exercises for Section 13.4

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

13.4.1 VS Consider the following computer output from a RCBD.

Source	DF	SS	MS	F	P
Factor	?	193.800	64.600	?	?
Block	3	464.218	154.739		
Error	?	?	4.464		
Total	15	698.190			

a. How many levels of the factor were used in this experiment?
b. How many blocks were used in this experiment?
c. Fill in the missing information. Use bounds for the *P*-value.
d. What conclusions would you draw if $\alpha = 0.05$? What would you conclude if $\alpha = 0.01$?

13.4.2 Exercise 13.2.4 introduced you to an experiment to investigate the potential effect of consuming chocolate on cardiovascular health. The experiment was conducted as a completely randomized design, and the exercise asked you to use the ANOVA to analyze the data and draw conclusions. Now assume that the experiment had been conducted as an RCBD with the subjects considered as blocks. Analyze the data using this assumption. What conclusions would you draw (using $\alpha = 0.05$) about the effect of the different types of chocolate on cardiovascular health? Would your conclusions change if $\alpha = 0.01$?

13.4.3 WP SS In "The Effect of Nozzle Design on the Stability and Performance of Turbulent Water Jets" (*Fire Safety Journal*, August 1981, Vol. 4(1), pp. 1–13), C. Theobald described an experiment in which a shape measurement was determined for several different nozzle types at different levels of jet efflux velocity. Interest in this experiment focuses primarily on nozzle type, and velocity is a nuisance factor. The data are as follows:

	Jet Efflux Velocity (m/s)					
Nozzle Type	**11.73**	**14.37**	**16.59**	**20.43**	**23.46**	**28.74**
1	0.78	0.80	0.81	0.75	0.77	0.78
2	0.85	0.85	0.92	0.86	0.81	0.83
3	0.93	0.92	0.95	0.89	0.89	0.83
4	1.14	0.97	0.98	0.88	0.86	0.83
5	0.97	0.86	0.78	0.76	0.76	0.75

a. Does nozzle type affect shape measurement? Compare the nozzles with box plots and the analysis of variance.
b. Use Fisher's LSD method to determine specific differences among the nozzles. Does a graph of the average (or standard deviation) of the shape measurements versus nozzle type assist with the conclusions?
c. Analyze the residuals from this experiment.

13.4.4 WP An article in *Quality Engineering* ["Designed Experiment to Stabilize Blood Glucose Levels" (1999–2000, Vol. 12, pp. 83–87)] described an experiment to minimize variations in blood glucose levels. The treatment was the exercise time on a Nordic Track cross-country skier (10 or 20 min). The experiment was blocked for time of day. The data were as follows:

Exercise (min)	Time of Day	Average Blood Glucose
10	pm	71.5
10	am	103.0
20	am	83.5
20	pm	126.0
10	am	125.5
10	pm	129.5
20	pm	95.0
20	am	93.0

a. Is there an effect of exercise time on the average blood glucose? Use $\alpha = 0.05$.
b. Find the *P*-value for the test in part (a).
c. Analyze the residuals from this experiment.

13.4.5 WP SS VS An article in the *American Industrial Hygiene Association Journal* ["A Rapid Method for the Determination of Arsenic Concentrations in Urine at Field Locations" (1976, Vol. 37(7), pp. 418–422)] described a field test for detecting the presence of arsenic in urine samples. The test has been proposed for use among forestry workers because of the increasing use of organic arsenics in that industry. The experiment compared the test as performed by both a trainee and an experienced trainer to an analysis at a remote laboratory. Four subjects were selected for testing and are considered as blocks. The response variable is arsenic content (in ppm) in the subject's urine. The data are as follows:

	Subject			
Test	**1**	**2**	**3**	**4**
Trainee	0.05	0.05	0.04	0.15
Trainer	0.05	0.05	0.04	0.17
Lab	0.04	0.04	0.03	0.10

a. Is there any difference in the arsenic test procedure?
b. Analyze the residuals from this experiment.

13.4.6 In *Design and Analysis of Experiments*, 9th edition (John Wiley & Sons, 2017), D. C. Montgomery described an experiment that determined the effect of four different types of tips in a hardness tester on the observed hardness of a metal alloy. Four specimens of the alloy were obtained, and each tip was tested once on each specimen, producing the following data:

	Specimen			
Type of Tip	**1**	**2**	**3**	**4**
1	9.3	9.4	9.6	10.0
2	9.4	9.3	9.8	9.9
3	9.2	9.4	9.5	9.7
4	9.7	9.6	10.0	10.2

a. Is there any difference in hardness measurements between the tips?
b. Use Fisher's LSD method to investigate specific differences between the tips.
c. Analyze the residuals from this experiment.

13.4.7 An experiment was conducted to investigate leaking current in a SOS MOSFETS device. The purpose of the experiment was to investigate how leakage current varies as the channel length changes. Four channel lengths were selected. For each channel length, five different widths were also used, and width is to be considered a nuisance factor. The data are as follows:

	Width				
Channel Length	**1**	**2**	**3**	**4**	**5**
1	0.7	0.8	0.8	0.9	1.0
2	0.8	0.8	0.9	0.9	1.0
3	0.9	1.0	1.7	2.0	4.0
4	1.0	1.5	2.0	3.0	20.0

a. Test the hypothesis that mean leakage voltage does not depend on the channel length using $\alpha = 0.05$.
b. Analyze the residuals from this experiment. Comment on the residual plots.
c. The observed leakage voltage for channel length 4 and width 5 was erroneously recorded. The correct observation is 4.0. Analyze the corrected data from this experiment. Is there evidence to conclude that mean leakage voltage increases with channel length?

13.4.8 An article in the *Food Technology Journal* (1956, Vol. 10, pp. 39–42) described a study on the protopectin content of tomatoes during storage. Four storage times were selected, and samples from nine lots of tomatoes were analyzed. The protopectin content (expressed as hydrochloric acid soluble fraction mg/kg) is in the table below.

a. The researchers in this study hypothesized that mean protopectin content would be different at different storage times. Can you confirm this hypothesis with a statistical test using $\alpha = 0.05$?
b. Find the *P*-value for the test in part (a).
c. Which specific storage times are different? Would you agree with the statement that protopectin content decreases as storage time increases?
d. Analyze the residuals from this experiment.

	Lot								
Storage Time	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0 days	1694.0	989.0	917.3	346.1	1260.0	965.6	1123.0	1106.0	1116.0
7 days	1802.0	1074.0	278.8	1375.0	544.0	672.2	818.0	406.8	461.6
14 days	1568.0	646.2	1820.0	1150.0	983.7	395.3	422.3	420.0	409.5
21 days	415.5	845.4	377.6	279.4	447.8	272.1	394.1	356.4	351.2

Supplemental Exercises for Chapter 13

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

13.S9 Consider the following computer output.

Source	DF	SS	MS	F	P
Factor	?	126.880	63.4401	?	?
Block	?	54.825	18.2751		
Error	6	?	2.7403		
Total	11	198.147			

a. How many levels of the factor were used in this experiment?
b. How many blocks were used?
c. Fill in the missing information. Use bounds for the *P*-value.
d. What conclusions would you draw if $\alpha = 0.05$? What if $\alpha = 0.01$?

13.S10 **WP** An article in *Lubrication Engineering* (December 1990) described the results of an experiment designed to investigate the effects of carbon material properties on the progression of blisters on carbon face seals. The carbon face seals are used extensively in equipment such as air turbine starters. Five different carbon materials were tested, and the surface roughness was measured. The data are as follows:

Carbon Material Type	**Surface Roughness**					
EC10	0.50	0.55	0.55	0.36		
EC10A	0.31	0.07	0.25	0.18	0.56	0.20
EC4	0.20	0.28	0.12			
EC1	0.10	0.16				

a. Does carbon material type have an effect on mean surface roughness? Use $\alpha = 0.05$.
b. Find the residuals for this experiment. Does a normal probability plot of the residuals indicate any problem with the normality assumption?
c. Plot the residuals versus $\hat{y}_{ij}$. Comment on the plot.
d. Find a 95% confidence interval on the difference between mean surface roughness for the EC10 and the EC1 carbon grades.
e. Apply the Fisher LSD method to this experiment. Summarize your conclusions regarding the effect of material type on surface roughness.

13.S11 **WP** An article in the *IEEE Transactions on Components, Hybrids, and Manufacturing Technology* [(1992, Vol. 15(2), pp. 146–153)] described an experiment in which the contact resistance of a brake-only relay was studied for three different materials (all were silver-based alloys). The data are as follows.

Alloy	**Contact Resistance**				
1	95	97	99	98	99
	99	99	94	95	98
2	104	102	102	105	99
	102	111	103	100	103
3	119	130	132	136	141
	172	145	150	144	135

a. Does the type of alloy affect mean contact resistance? Use $\alpha = 0.01$.
b. Use Fisher's LSD method to determine which means differ.
c. Find a 99% confidence interval on the mean contact resistance for alloy 3.
d. Analyze the residuals for this experiment.

13.S12 An article in the *Journal of Quality Technology* [(1982, Vol. 14(2), pp. 80–89)] described an experiment in which three different methods of preparing fish were evaluated on the basis of sensory criteria, and a quality score was assigned. Assume that these methods have been randomly selected from a large population of preparation methods. The data are in the following table:

Method	**Score**			
1	24.4	23.2	25.0	19.7
	22.2	24.4	23.8	18.0
2	22.1	19.5	17.3	19.7
	22.3	23.2	21.4	22.6
3	23.3	22.8	22.4	23.7
	20.4	23.5	20.8	24.1

a. Is there any difference in preparation methods? Use $\alpha = 0.05$.

b. Calculate the P-value for the F-statistic in part (a).

c. Analyze the residuals from this experiment and comment on model adequacy.

d. Estimate the components of variance.

13.S13 WP SS An article in the *Journal of Agricultural Engineering Research* (1992, Vol. 52, pp. 53–76) described an experiment to investigate the effect of drying temperature of wheat grain on baking quality bread. Three temperature levels were used, and the response variable measured was the volume of the loaf of bread produced. The data are as follows:

Temperature (°C)	Volume (cc)				
70.0	1245	1235	1285	1245	1235
75.0	1235	1240	1200	1220	1210
80.0	1225	1200	1170	1155	1095

a. Does drying temperature affect mean bread volume? Use $\alpha = 0.01$.

b. Find the P-value for this test.

c. Use the Fisher LSD method to determine which means are different.

d. Analyze the residuals from this experiment and comment on model adequacy.

13.S14 An article in *Nature Genetics* ["Treatment-Specific Changes in Gene Expression Discriminate In Vivo Drug Response in Human Leukemia Cells" (2003, Vol. 34(1), pp. 85–90)] reported the results of a study of gene expression as a function of different treatments for leukemia. Three treatment groups are mercaptopurine (MP) only, low-dose methotrexate (LDMTX) and MP, and high-dose methotrexate (HDMTX) and MP. Each group contained 10 subjects. The responses from a specific gene are shown below.

Treatments	Gene Expression				
MP ONLY	334.5	31.6	701	41.2	61.2
	69.6	67.5	66.6	120.7	881.9
MP + HDMTX	919.4	404.2	1024.8	54.1	62.8
	671.6	882.1	354.2	321.9	91.1
MP + LDMTX	108.4	26.1	240.8	191.1	69.7
	242.8	62.7	396.9	23.6	290.4

a. Check the normality of the data. Can you assume that these samples are from normal populations?

b. Take the logarithm of the raw data and check the normality of the transformed data. Is there evidence to support the claim that the treatment means differ for the transformed data? Use $\alpha = 0.1$.

c. Analyze the residuals from the transformed data and comment on model adequacy.

13.S15 WP An article in *Communications of the ACM* [(1987, Vol. 30(5), pp. 53–76] reported on a study of different algorithms for estimating software development costs. Six algorithms were applied to eight software development projects (nuisance factor) and the percent error in estimating the development cost was observed. The data are in the table below.

a. Do the algorithms differ in mean cost estimation accuracy? Use $\alpha = 0.05$.

b. Analyze the residuals from this experiment.

c. Which algorithm would you recommend for use in practice?

13.S16 An article in the *Journal of Hazardous Materials* ["Toxicity Assessment from Electro-Coagulation Treated-Textile Dye Waste Waters by Bioassays" (2009, Vol. 172(1), pp. 330–337)] discussed a study of pollutant removal from textile dyeing waste water with an electro-coagulation technique. Chemical oxygen demand (COD) (a common measure of water pollution) was used as the response, and three different values for electrolysis time were considered. The following data were extracted from a larger study. Suppose that a randomized complete block experiment was conducted with three blocks based on initial pH values.

Electrolysis time (min)	Initial pH		
	3	7	11
15	77.1	75.2	42.2
30	80.1	76.8	45.0
45	82.8	75.2	46.8

a. Is there an effect of electrolysis time at $\alpha = 0.05$? Calculate the P-value.

b. Analyze the residuals from the experiment.

c. Calculate a 95% confidence interval on mean COD removal when the electrolysis time is 15 minutes.

d. Perform an ANOVA assuming that all data are collected at a single pH value. Comment on differences from part (a).

Algorithm	Project							
	1	2	3	4	5	6	7	8
1(SLIM)	1244	21	82	2221	905	839	527	122
2(COCOMO-A)	281	129	396	1306	336	910	473	199
3(COCOMO-R)	220	84	458	543	300	794	488	142
4(COCOMO-C)	225	83	425	552	291	826	509	153
5(FUNCTION POINTS)	19	11	−34	121	15	103	87	−17
6(ESTIMALS)	−20	35	−53	170	104	199	142	41

Chapter 14 Exercises

Exercises for Section 14.3

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

14.3.1 WP An article in *Industrial Quality Control* (1956, pp. 5–8) describes an experiment to investigate the effect of two factors (glass type and phosphor type) on the brightness of a television tube. The response variable measured is the current (in microamps) necessary to obtain a specified brightness level. The data are shown in the following table:

	Phosphor Type		
Glass Type	**1**	**2**	**3**
1	280	300	290
	290	310	285
	285	295	290
2	230	260	220
	235	240	225
	240	235	230

a. State the hypotheses of interest in this experiment.
b. Test the hypotheses in part (a) and draw conclusions using the analysis of variance with $\alpha = 0.05$.
c. Analyze the residuals from this experiment.

14.3.2 An engineer suspects that the surface finish of metal parts is influenced by the type of paint used and the drying time. He selected three drying times—20, 25, and 30 minutes—and used two types of paint. Three parts are tested with each combination of paint type and drying time. The data are as follows:

	Drying Time (min)		
Paint	**20**	**25**	**30**
1	74	73	78
	64	61	85
	50	44	92
2	92	98	66
	86	73	45
	68	88	85

a. State the hypotheses of interest in this experiment.
b. Test the hypotheses in part (a) and draw conclusions using the analysis of variance with $\alpha = 0.05$.
c. Analyze the residuals from this experiment.

14.3.3 An article in *Technometrics* ["Exact Analysis of Means with Unequal Variances" (2002, Vol. 44, pp. 152–160)] described the technique of the analysis of means (ANOM) and presented the results of an experiment on insulation. Four insulation types were tested at three different temperatures. The data are as follows:

	Temperature (°F)					
Insulation	**1**		**2**		**3**	
1	6.6	4	4.5	2.2	2.3	0.9
	2.7	6.2	5.5	2.7	5.6	4.9
	6	5	4.8	5.8	2.2	3.4
2	3	3.2	3	1.5	1.3	3.3
	2.1	4.1	2.5	2.6	0.5	1.1
	5.9	2.5	0.4	3.5	1.7	0.1
3	5.7	4.4	8.9	7.7	2.6	9.9
	3.2	3.2	7	7.3	11.5	10.5
	5.3	9.7	8	2.2	3.4	6.7
4	7	8.9	12	9.7	8.3	8
	7.3	9	8.5	10.8	10.4	9.7
	8.6	11.3	7.9	7.3	10.6	7.4

a. Write a model for this experiment.
b. Test the appropriate hypotheses and draw conclusions using the analysis of variance with $\alpha = 0.05$.
c. Graphically analyze the interaction.
d. Analyze the residuals from the experiment.
e. Use Fisher's LSD method to investigate the differences between mean effects of insulation type. Use $\alpha = 0.05$.

14.3.4 An experiment was conducted to determine whether either firing temperature or furnace position affects the baked density of a carbon anode. The data are as follows:

	Temperature (°C)		
Position	**800**	**825**	**850**
1	570	1063	565
	565	1080	510
	583	1043	590
2	528	988	526
	547	1026	538
	521	1004	532

a. State the hypotheses of interest.
b. Test the hypotheses in part (a) using the analysis of variance with $\alpha = 0.05$. What are your conclusions?
c. Analyze the residuals from this experiment.
d. Using Fisher's LSD method, investigate the differences between the mean baked anode density at the three different levels of temperature. Use $\alpha = 0.05$.

14.3.5 WP SS An article in the *IEEE Transactions on Electron Devices* (1986, Vol. 33, p. 1754) described a study on the effects of two variables—polysilicon doping and anneal conditions (time and temperature)—on the base current of a bipolar transistor. The data from this experiment follow.

a. Is there any evidence to support the claim that either polysilicon doping level or anneal conditions affect base current? Do these variables interact? Use $\alpha = 0.05$.

b. Graphically analyze the interaction.

c. Analyze the residuals from this experiment.

d. Use Fisher's LSD method to isolate the effects of anneal conditions on base current, with $\alpha = 0.05$.

	Anneal (temperature/time)				
Polysilicon doping	900 60	900 180	950 60	1000 15	1000 30
1×10^{20}	4.40	8.30	10.15	10.29	11.01
	4.60	8.90	10.20	10.30	10.58
2×10^{20}	3.20	7.81	9.38	10.19	10.81
	3.50	7.75	10.02	10.10	10.60

14.3.6 An article in the *Journal of Testing and Evaluation* (1988, Vol. 16, pp. 508–515) investigated the effects of cyclic loading frequency and environment conditions on fatigue crack growth at a constant 22 MPa stress for a particular material. The data follow. The response variable is fatigue crack growth rate.

	Environment		
Frequency	**Air**	**H_2O**	**Salt H_2O**
10	2.29	2.06	1.90
	2.47	2.05	1.93
	2.48	2.23	1.75
	2.12	2.03	2.06
1	2.65	3.20	3.10
	2.68	3.18	3.24
	2.06	3.96	3.98
	2.38	3.64	3.24
0.1	2.24	11.00	9.96
	2.71	11.00	10.01
	2.81	9.06	9.36
	2.08	11.30	10.40

a. Is there indication that either factor affects crack growth rate? Is there any indication of interaction? Use $\alpha = 0.05$.

b. Analyze the residuals from this experiment.

c. Repeat the analysis in part (a) using $ln(y)$ as the response. Analyze the residuals from this new response variable and comment on the results.

14.3.7 VS An article in *Bioresource Technology* ["Quantitative Response of Cell Growth and Tuber Polysaccharides Biosynthesis by Medicinal Mushroom Chinese Truffle Tuber Sinense to Metal Ion in Culture Medium" (2008, Vol. 99(16), pp. 7606–7615)] described an experiment to investigate the effect of metal ion concentration to the production of extracellular polysaccharides (EPS). It is suspected that Mg^{2+} and K^+ (in millimolars) are related to EPS. The data from a full factorial design follow.

Run	**Mg^{2+} (mM)**	**K^+ (mM)**	**EPS (g/L)**
1	40	5	3.88
2	50	15	4.23
3	40	10	4.67
4	30	5	5.86
5	50	10	4.50
6	50	5	3.62
7	30	15	3.84
8	40	15	3.25
9	30	10	4.18

a. State the hypotheses of interest.

b. Test the hypotheses with $\alpha = 0.5$.

c. Analyze the residuals and plot residuals versus the predicted production.

Exercises for Section 14.4

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

14.4.1 WP SS VS The percentage of hardwood concentration in raw pulp, the freeness, and the cooking time of the pulp are being investigated for their effects on the strength of paper. The data from a three-factor factorial experiment are shown in the following table.

a. Analyze the data using the analysis of variance assuming that all factors are fixed. Use $\alpha = 0.05$.

b. Compute approximate P-values for the F-ratios in part (a).

c. The residuals are found from $e_{ijkl} = y_{ijkl} - \bar{y}_{ijk\cdot}$. Graphically analyze the residuals from this experiment.

	Cooking Time 1.5 hours			Cooking Time 2.0 hours		
	Freeness			Freeness		
Hardwood Concentration %	350	500	650	350	500	650
10	96.6	97.7	99.4	98.4	99.6	100.6
	96.0	96.0	99.8	98.6	100.4	100.9
15	98.5	96.0	98.4	97.5	98.7	99.6
	97.2	96.9	97.6	98.1	96.0	99.0
20	97.5	95.6	97.4	97.6	97.0	98.5
	96.6	96.2	98.1	98.4	97.8	99.8

14.4.2 The quality control department of a fabric finishing plant is studying the effects of several factors on dyeing for a blended cotton/synthetic cloth used to manufacture shirts. Three operators, three cycle times, and two temperatures were selected, and three small specimens of cloth were dyed under each set of conditions. The finished cloth was compared to a standard, and a numerical score was assigned. The results are shown in the following table.

a. State and test the appropriate hypotheses using the analysis of variance with $\alpha = 0.05$.

	Temperature					
	300°			350°		
	Operator			Operator		
Cycle Time	1	2	3	1	2	3
40	23	27	31	24	38	34
	24	28	32	23	36	36
	25	26	28	28	35	39
50	36	34	33	37	34	34
	35	38	34	39	38	36
	36	39	35	35	36	31
60	28	35	26	26	36	28
	24	35	27	29	37	26
	27	34	25	25	34	34

b. The residuals may be obtained from $e_{ijkl} = y_{ijkl} - \bar{y}_{ijk\cdot}$. Graphically analyze the residuals from this experiment.

Exercises for Section 14.5

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

14.5.1 VS Four factors are thought to influence the taste of a soft-drink beverage: type of sweetener (A), ratio of syrup to water (B), carbonation level (C), and temperature (D). Each factor can be run at two levels, producing a 2^4 design. At each run in the design, samples of the beverage are given to a test panel consisting of 20 people. Each tester assigns the beverage a point score from 1 to 10. Total score is the response variable, and the objective is to find a formulation that maximizes total score. Two replicates of this design are run, and the results are shown in the table. Analyze the data and draw conclusions. Use $\alpha = 0.05$ in the statistical tests.

	Replicate	
Treatment Combination	I	II
(1)	159	163
a	168	175
b	158	163
ab	166	168
c	175	178
ac	179	183
bc	173	168
abc	179	182
d	164	159
ad	187	189
bd	163	159
abd	185	191
cd	168	174
acd	197	199
bcd	170	174
$abcd$	194	198

14.5.2 WP An engineer is interested in the effect of cutting speed (A), metal hardness (B), and cutting angle (C) on the life of a cutting tool. Two levels of each factor are chosen, and two replicates of a 2^3 factorial design are run. The tool life data (in hours) are shown in the following table.

Treatment Combination	Replicate I	Replicate II
(1)	221	311
a	325	435
b	354	348
ab	552	472
c	440	453
ac	406	377
bc	605	500
abc	392	419

a. Analyze the data from this experiment.
b. Find an appropriate regression model that explains tool life in terms of the variables used in the experiment.
c. Analyze the residuals from this experiment.

14.5.3 WP SS An article in *IEEE Transactions on Semiconductor Manufacturing* (1992, Vol. 5, pp. 214–222) described an experiment to investigate the surface charge on a silicon wafer. The factors thought to influence induced surface charge are cleaning method (spin rinse dry or SRD and spin dry or SD) and the position on the wafer where the charge was measured. The surface charge ($\times 10^{11}$ q/cm^3) response data follow:

Cleaning Method	Test Position L	Test Position R
SD	1.66	1.84
	1.90	1.84
	1.92	1.62
SRD	−4.21	−7.58
	−1.35	−2.20
	−2.08	−5.36

a. Estimate the factor effects.
b. Which factors appear important? Use $\alpha = 0.05$.
c. Analyze the residuals from this experiment.

14.5.4 An article in *Talanta* (2005, Vol. 65, pp. 895–899) presented a 2^3 factorial design (with two replicates) to find lead level by using flame atomic absorption spectrometry (FAAS). The data are in the table at the top of the next column.

Run	Factors ST	Factors pH	Factors RC	Lead Recovery (%) R1	Lead Recovery (%) R2
1	−	−	−	39.8	42.1
2	+	−	−	51.3	48
3	−	+	−	57.9	58.1
4	+	+	−	78.9	85.9
5	−	−	+	78.9	84.2
6	+	−	+	84.2	84.2
7	−	+	+	94.4	90.9
8	+	+	+	94.7	105.3

The factors and levels are in the following table.

Factor	Low (−)	High (+)
Reagent concentration (RC) (mol 1^{-1})	5×10^{-6}	5×10^{-5}
pH	6.0	8.0
Shaking time (ST) (min)	10	30

a. Construct a normal probability plot of the effect estimates. Which effects appear to be large?
b. Conduct an analysis of variance to confirm your findings for part (a).
c. Analyze the residuals from this experiment. Are there any problems with model adequacy?

14.5.5 An article in the *Journal of Construction Engineering and Management* ["Analysis of Earth-Moving Systems Using Discrete-Event Simulation" (1995, Vol. 121(4), pp. 388–396)] considered a replicated two-level factorial experiment to study the factors most important to output in an earth-moving system. Handle the experiment as four replicates of a 2^4 factorial design with response equal to production rate (m^3/h). The data are shown in the following table.

a. Estimate the factor effects. Based on a normal probability plot of the effect estimates, identify a model for the data from this experiment.
b. Conduct an ANOVA based on the model identified in part (a). What are the conclusions?
c. Analyze the residuals and plot residuals versus the predicted production.
d. Comment on model adequacy.

Row	Number of Trucks	Passes per Load	Load-pass Time	Travel Time	Production (m^3/h) 1	2	3	4
1	−	−	−	−	179.6	179.8	176.3	173.1
2	+	−	−	−	373.1	375.9	372.4	361.1
3	−	+	−	−	153.2	153.6	150.8	148.6
4	+	+	−	−	226.1	220.0	225.7	218.5
5	−	−	+	−	156.9	155.4	154.2	152.2
6	+	−	+	−	242.0	233.5	242.3	233.6
7	−	+	+	−	122.7	119.6	120.9	118.6
8	+	+	+	−	135.7	130.9	135.5	131.6
9	−	−	−	+	44.2	44.0	43.5	43.6
10	+	−	−	+	124.2	123.3	122.8	121.6
11	−	+	−	+	42.0	42.4	42.5	41.0
12	+	+	−	+	116.3	117.3	115.6	114.7
13	−	−	+	+	42.1	42.6	42.8	42.9
14	+	−	+	+	119.1	119.5	116.9	117.2
15	−	+	+	+	39.6	39.7	39.5	39.2
16	+	+	+	+	107.0	105.3	104.2	103.0

Level	Number of trucks	Passes per load	Load pass time	Travel time
−1	2	4	12 s	100 s
1	6	7	22 s	800 s

Exercises for Section 14.6

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

14.6.1 **WP** Consider the data from Exercise 14.5.2. Suppose that the data from the second replicate were not available. Analyze the data from replicate I only and comment on your findings.

14.6.2 The following data represent a single replicate of a 2^5 design that is used in an experiment to study the compressive strength of concrete. The factors are mix (A), time (B), laboratory (C), temperature (D), and drying time (E).

$(1) = 700$ $e = 800$
$a = 900$ $ae = 1200$
$b = 3400$ $be = 3500$
$ab = 5500$ $abe = 6200$
$c = 600$ $ce = 600$
$ac = 1000$ $ace = 1200$
$bc = 3000$ $bce = 3006$
$abc = 5300$ $abce = 5500$
$d = 1000$ $de = 1900$
$ad = 1100$ $ade = 1500$
$bd = 3000$ $bde = 4000$
$abd = 6100$ $abde = 6500$
$cd = 800$ $cde = 1500$
$acd = 1100$ $acde = 2000$
$bcd = 3300$ $bcde = 3400$
$abcd = 6000$ $abcde = 6800$

a. Estimate the factor effects.
b. Which effects appear important? Use a normal probability plot.
c. If it is desirable to maximize the strength, in which direction would you adjust the process variables?
d. Analyze the residuals from this experiment.

14.6.3 WP An experiment has run a single replicate of a 2^4 design and calculated the following factor effects:

$A = 80.25$	$AB = 53.25$	$ABC = -2.95$
$B = -65.50$	$AC = 11.00$	$ABD = -8.00$
$C = -9.25$	$AD = 9.75$	$ACD = 10.25$
$D = -20.50$	$BC = 18.36$	$BCD = -7.95$
	$BD = 15.10$	$ABCD = -6.25$
	$CD = -1.25$	

a. Construct a normal probability plot of the effects.
b. Identify a tentative model, based on the plot of effects in part (a).
c. Estimate the regression coefficients in this model, assuming that $\bar{y} = 400$.

14.6.4 VS An experiment described by M. G. Natrella in the National Bureau of Standards' *Handbook of Experimental Statistics* (1963, No. 91) involves flame-testing fabrics after applying fire-retardant treatments. The four factors considered are type of fabric (A), type of fire-retardant treatment (B), laundering condition (C—the low level is no laundering, the high level is after one laundering), and method of conducting the flame test (D). All factors are run at two levels, and the response variable is the inches of fabric burned on a standard-size test sample. The data are:

$(1) = 42$	$d = 40$
$a = 31$	$ad = 30$
$b = 45$	$bd = 50$
$ab = 29$	$abd = 25$
$c = 39$	$cd = 40$
$ac = 28$	$acd = 25$
$bc = 46$	$bcd = 50$
$abc = 32$	$abcd = 23$

a. Estimate the effects and prepare a normal plot of the effects.
b. Construct an analysis of variance table based on the model tentatively identified in part (a).
c. Construct a normal probability plot of the residuals and comment on the results.

14.6.5 SS Consider the following computer output for one replicate of a 2^4 factorial experiment.

a. What effects are used to estimate error?
b. Calculate the entries marked with "?" in the output.

Estimated Effects and Coefficients

Term	Effect	Coef	SE Coef	t	P
Constant		35.250	?	39.26	0.000
A	2.250	?	?	?	?
B	24.750	12.375	?	13.78	0.000
C	1.000	0.500	?	0.56	0.602
D	10.750	5.375	?	5.99	0.002
A*B	−10.500	−5.250	?	−5.85	0.002
A*C	4.250	2.125	?	2.37	0.064
A*D	−5.000	−2.500	?	−2.78	0.039
B*C	5.250	2.625	?	2.92	0.033
B*D	4.000	2.000	?	2.23	0.076
C*D	−0.750	−0.375	?	−0.42	0.694

S = 3.59166

Analysis of Variance

Source	DF	SS	MS	F	P
A	?	?	?	?	0.266
B	1	2450.25	2450.25	189.94	0.000
C	1	4.00	4.00	0.31	0.602
D	1	462.25	462.25	35.83	0.002
AB	1	441.00	441.00	34.19	0.002
AC	1	72.25	72.25	5.60	0.064
AD	1	100.00	100.00	7.75	0.039
BC	1	110.25	110.25	8.55	0.033
BD	1	64.00	64.00	4.96	0.076
CD	1	2.25	2.25	0.17	0.694
Residual Error	?	64.50	?		
Total	?	3791.00			

14.6.6 An article in *Bioresource Technology* ["Influence of Vegetable Oils Fatty-Acid Composition on Biodiesel Optimization" (2011, Vol. 102(2), pp. 1059–1065)] described an experiment to analyze the influence of the fatty-acid composition on biodiesel. Factors were the concentration of catalyst, amount of methanol, reaction temperature and time, and the design included three center points. Maize oil methyl ester (MME) was recorded as the response. Data follow.

Run	Temperature (°C)	Time (min)	Catalyst (wt.%)	Methanol to oil molar ratio	MME (wt.%)
1	45	40	0.8	5.4	88.30
2	25	40	1.2	5.4	90.50
3	45	10	0.8	4.2	77.96
4	25	10	1.2	5.4	85.59
5	45	40	1.2	5.4	97.14
6	45	10	1.2	4.2	90.64
7	45	40	1.2	4.2	89.86
8	25	40	0.8	4.2	82.35
9	25	10	0.8	5.4	80.31
10	25	40	0.8	5.4	85.51
11	25	10	0.8	4.2	76.21
12	45	40	0.8	4.2	86.86
13	25	10	1.2	4.2	86.35
14	45	10	0.8	5.4	84.58
15	25	40	1.2	4.2	89.37
16	45	10	1.2	5.4	90.51
17	35	25	1	4.8	91.40
18	35	25	1	4.8	91.96
19	35	25	1	4.8	91.07

a. Identify the important effects from a normal probability plot.
b. Compare the results in the previous part with results that use an error term based on the center points.
c. Test for curvature.
d. Analyze the residuals from the model.

14.6.7 An article in *Analytica Chimica Acta* ["Design-of-Experiment Optimization of Exhaled Breath Condensate Analysis Using a Miniature Differential Mobility Spectrometer (DMS)" (2008, Vol. 628(2), pp. 155–161)] examined four parameters that affect the sensitivity and detection of the analytical instruments used to measure clinical samples. They optimized the sensor function using exhaled breath condensate (EBC) samples spiked with acetone, a known clinical biomarker in breath. The following table shows the results for a single replicate of a 2^4 factorial experiment for one of the outputs, the average amplitude of acetone peak over three repetitions.

Configuration	A	B	C	D	*y*
1	+	+	+	+	0.12
2	+	+	+	−	0.1193
3	+	+	−	+	0.1196
4	+	+	−	−	0.1192
5	+	−	+	+	0.1186
6	+	−	+	−	0.1188
7	+	−	−	+	0.1191
8	+	−	−	−	0.1186
9	−	+	+	+	0.121
10	−	+	+	−	0.1195
11	−	+	−	+	0.1196
12	−	+	−	−	0.1191
13	−	−	+	+	0.1192
14	−	−	+	−	0.1194
15	−	−	−	+	0.1188
16	−	−	−	−	0.1188

The factors and levels are shown in the following table.

A	RF voltage of the DMS sensor (1200 or 1400 V)
B	Nitrogen carrier gas flow rate (250 or 500 mL min^{-1})
C	Solid phase microextraction (SPME) filter type (polyacrylate or PDMS–DVB)
D	GC cooling profile (cryogenic and noncryogenic)

a. Estimate the factor effects and use a normal probability plot of the effects. Identify which effects appear to be large, and identify a model for the data from this experiment.
b. Conduct an ANOVA based on the model identified in part (a). What are your conclusions?
c. Analyze the residuals from this experiment. Are there any problems with model adequacy?
d. Project the design in this problem into a 2^r design for $r < 4$ in the important factors. Sketch the design and show the average and range of yields at each run. Does this sketch aid in data representation?

Exercises for Section 14.7

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

14.7.1 WP SS Consider a 2^2 factorial experiment with four center points. The data are (1) = 21, a = 125, b = 154, ab = 352, and the responses at the center point are 92, 130, 98, and 152. Compute an ANOVA with the sum of squares for curvature and conduct an F-test for curvature. Use $\alpha = 0.05$.

14.7.2 WP Consider the experiment in Exercise 14.6.2. Suppose that a center point with five replicates is added to the factorial runs and the responses are 2800, 5600, 4500, 5400, and 3600. Compute an ANOVA with the sum of squares for curvature and conduct an F-test for curvature. Use $\alpha = 0.05$.

Exercises for Section 14.8

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

14.8.1 WP Consider the data from the first replicate of Exercise 14.5.2.
a. Construct a design with two blocks of four observations each with *ABC* confounded.
b. Analyze the data.

14.8.2 Consider the data from Exercise 14.6.4.
a. Construct the design that would have been used to run this experiment in two blocks of eight runs each.
b. Analyze the data and draw conclusions.

14.8.3 Construct a 2^5 design in two blocks. Select the *ABCDE* interaction to be confounded with blocks.

14.8.4 Consider the data from the first replicate of Exercise 14.5.1, assuming that four blocks are required. Confound *ABD* and *ABC* (and consequently *CD*) with blocks.

a. Construct a design with four blocks of four observations each.

b. Analyze the data.

14.8.5 SS An article in *Quality Engineering* ["Designed Experiment to Stabilize Blood Glucose Levels" (1999–2000, Vol. 12, pp. 83–87)] reported on an experiment to minimize variations in blood glucose levels. The factors were volume of juice intake before exercise (4 or 8 oz), amount of exercise on a Nordic Track cross-country skier (10 or 20 min), and delay between the time of juice intake (0 or 20 min) and the beginning of the exercise period. The experiment was blocked for time of day. The data follow.

a. What effects are confounded with blocks? Comment on any concerns with the confounding in this design.

b. Analyze the data and draw conclusions.

Run	Juice (oz)	Exercise (min)	Delay (min)	Time of Day	Average Blood Glucose
1	4	10	0	pm	71.5
2	8	10	0	am	103
3	4	20	0	am	83.5
4	8	20	0	pm	126
5	4	10	20	am	125.5
6	8	10	20	pm	129.5
7	4	20	20	pm	95
8	8	20	20	am	93

14.8.6 Construct a 2^6 design in four blocks. Show that a design that confounds three of the four-factor interactions with blocks is the best possible blocking arrangement.

14.8.7 An article in *Advanced Semiconductor Manufacturing Conference (ASMC)* (May 2004, pp. 325–29) stated that dispatching rules and rework strategies are two major operational elements that impact productivity in a semiconductor fabrication plant (fab). A four-factor experiment was conducted to determine the effect of dispatching rule time (5 or 10 min), rework delay (0 or 15 min), fab temperature (60 or 80°F), and rework levels (level 0 or level 1) on key fab performance measures. The performance measure that was analyzed was the average cycle time. The experiment was blocked for the fab temperature. Data modified from the original study are in the following table.

Run	Dispatching Rule Time (min)	Rework Delay (min)	Rework Level	Fab Temperature (°F)	Average Cycle TimeRun (min)
1	5	0	0	60	218
2	10	0	0	80	256.5
3	5	0	1	80	231
4	10	0	1	60	302.5
5	5	15	0	80	298.5
6	10	15	0	60	314
7	5	15	1	60	249
8	10	15	1	80	241

a. What effects are confounded with blocks? Do you find any concerns with confounding in this design? If so, comment on it.

b. Analyze the data and draw conclusions.

Exercises for Section 14.9

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

14.9.1 WP An article by L. B. Hare ["In the Soup: A Case Study to Identify Contributors to Filling Variability," *Journal of Quality Technology* (1988, Vol. 20, pp. 36–43)] described a factorial experiment used to study filling variability of dry soup mix packages. The factors are A = number of mixing ports through which the vegetable oil was added (1, 2), B = temperature surrounding the mixer (cooled, ambient), C = mixing time (60, 80 sec), D = batch weight (1500, 2000 lb), and E = number of days of delay between mixing and packaging (1, 7). Between 125 and 150 packages of soup were sampled over an 8-hour period for each run in the design, and the standard deviation of package weight was used as the response variable. The design and resulting data follow.

a. What is the generator for this design?

b. What is the resolution of this design?

c. Estimate the factor effects. Which effects are large?

d. Does a residual analysis indicate any problems with the underlying assumptions?

e. Draw conclusions about this filling process.

Std Order	*A* Mixer Ports	*B* Temp	*C* Time	*D* Batch Weight	*E* Delay	*y* Std Dev
1	−	−	−	−	−	1.13
2	+	−	−	−	+	1.25
3	−	+	−	−	+	0.97
4	+	+	−	−	−	1.70
5	−	−	+	−	+	1.47
6	+	−	+	−	−	1.28
7	−	+	+	−	−	1.18
8	+	+	+	−	+	0.98
9	−	−	−	+	+	0.78
10	+	−	−	+	−	1.36

11	−	+	−	+	−	1.85
12	+	+	−	+	+	0.62
13	−	−	+	+	−	1.09
14	+	−	+	+	+	1.10
15	−	+	+	+	+	0.76
16	+	+	+	+	−	2.10

14.9.2 An article in *Quality Engineering* ["A Comparison of Multi-Response Optimization: Sensitivity to Parameter Selection" (1999, Vol. 11, pp. 405–415)] conducted a half replicate of a 2^5 factorial design to optimize the retort process of beef stew MREs, a military ration. The design factors are x_1 = sauce viscosity, x_2 = residual gas, x_3 = solid/liquid ratio, x_4 = net weight, x_5 = rotation speed. The response variable is the heating rate index, a measure of heat penetration, and there are two replicates.

						Heating Rate Index	
Run	x_1	x_2	x_3	x_4	x_5	**I**	**II**
1	−1	−1	−1	−1	1	8.46	9.61
2	1	−1	−1	−1	−1	15.68	14.68
3	−1	1	−1	−1	−1	14.94	13.09
4	1	1	−1	−1	1	12.52	12.71
5	−1	−1	1	−1	−1	17.0	16.36
6	1	−1	1	−1	1	11.44	11.83
7	−1	1	1	−1	1	10.45	9.22
8	1	1	1	−1	−1	19.73	16.94
9	−1	−1	−1	1	−1	17.37	16.36
10	1	−1	−1	1	1	14.98	11.93
11	−1	1	−1	1	1	8.40	8.16
12	1	1	−1	1	−1	19.08	15.40
13	−1	−1	1	1	1	13.07	10.55
14	1	−1	1	1	−1	18.57	20.53
15	−1	1	1	1	−1	20.59	21.19
16	1	1	1	1	1	14.03	11.31

a. Estimate the factor effects. Based on a normal probability plot of the effect estimates, identify a model for the data from this experiment.

b. Conduct an ANOVA based on the model identified in part (a). What are your conclusions?

c. Analyze the residuals and comment on model adequacy.

d. Find a regression model to predict yield in terms of the coded factor levels.

e. This experiment was replicated, so an ANOVA could have been conducted without using a normal plot of the effects to tentatively identify a model. What model would be appropriate? Use the ANOVA to analyze this model and compare the results with those obtained from the normal probability plot approach.

14.9.3 WP SS R. D. Snee ("Experimenting with a Large Number of Variables," in *Experiments in Industry: Design, Analysis and Interpretation of Results*, Snee, Hare, and Trout, eds., ASQC, 1985) described an experiment in which a 2^{5-1} design with $I = ABCDE$ was used to investigate the effects of five factors on the color of a chemical product. The factors are A = solvent/reactant, B = catalyst/reactant, C = temperature, D = reactant purity, and E = reactant pH. The results obtained are as follows:

$e = -0.63$	$d = 6.79$
$a = 2.51$	$ade = 6.47$
$b = -2.68$	$bde = 3.45$
$abe = 1.66$	$abd = 5.68$
$c = 2.06$	$cde = 5.22$
$ace = 1.22$	$acd = 4.38$
$bce = -2.09$	$bcd = 4.30$
$abc = 1.93$	$abcde = 4.05$

a. Prepare a normal probability plot of the effects. Which factors are active?

b. Calculate the residuals. Construct a normal probability plot of the residuals and plot the residuals versus the fitted values. Comment on the plots.

c. If any factors are negligible, collapse the 2^{5-1} design into a full factorial in the active factors. Comment on the resulting design, and interpret the results.

Exercises for Section 14.10

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

14.10.1 WP An article in *Industrial and Engineering Chemistry* ["More on Planning Experiments to Increase Research Efficiency" (1970, pp. 60–65)] uses a 2^{5-2} design to investigate the effect on process yield of A = condensation temperature, B = amount of material 1, C = solvent volume, D = condensation time, and E = amount of material 2. The results obtained are as follows:

$ae = 23.2$	$cd = 23.8$
$ab = 15.5$	$ace = 23.4$
$ad = 16.9$	$bde = 16.8$
$bc = 16.2$	$abcde = 18.1$

a. Verify that the design generators used were $I = ACE$ and $I = BDE$.
b. Write down the complete defining relation and the aliases from the design.
c. Estimate the main effects.
d. Prepare an analysis of variance table. Verify that the AB and AD interactions are available to use as error.
e. Plot the residuals versus the fitted values. Also construct a normal probability plot of the residuals. Comment on the results.

14.10.2 WP Suppose that in Exercise 14.6.2 only a $\frac{1}{4}$ fraction of the 2^5 design could be run. Construct the design and analyze the data that are obtained by selecting only the response for the eight runs in your design.

14.10.3 WP SS For each of the following designs, write down the aliases, assuming that only main effects and two factor interactions are of interest.

a. 2^{6-3}_{III} **b.** 2^{8-4}_{IV}

14.10.4 WP Consider the 2^{6-2} design in Table 14.26.

a. Suppose that after analyzing the original data, we find that factors C and E can be dropped. What type of 2^k design is left in the remaining variables?
b. Suppose that after the original data analysis, we find that factors D and F can be dropped. What type of 2^k design is left in the remaining variables? Compare the results with part (a). Can you explain why the answers are different?

14.10.5 An article in the *Journal of Marketing Research* (1973, Vol. 10(3), pp. 270–276) presented a 2^{7-4} fractional factorial design to conduct marketing research:

Runs	*A*	*B*	*C*	*D*	*E*	*F*	*G*	Sales for a 6-Week Period (in $1000)
1	−1	−1	−1	1	1	1	−1	8.7
2	1	−1	−1	−1	−1	1	1	15.7
3	−1	1	−1	−1	1	−1	1	9.7
4	1	1	−1	1	−1	−1	−1	11.3
5	−1	−1	1	1	−1	−1	1	14.7
6	1	−1	1	−1	1	−1	−1	22.3
7	−1	1	1	−1	−1	1	−1	16.1
8	1	1	1	1	1	1	1	22.1

The factors and levels are shown in the following table.

	Factor	−1	+1
A	Television advertising	No advertising	Advertising
B	Billboard advertising	No advertising	Advertising
C	Newspaper advertising	No advertising	Advertising
D	Candy wrapper design	Conservative design	Flashy design
E	Display design	Normal shelf display	Special aisle display
F	Free sample program	No free samples	Free samples
G	Size of candy bar	1 oz bar	2 1/2 oz bar

a. Write down the alias relationships.
b. Estimate the main effects.
c. Prepare a normal probability plot for the effects and interpret the results.

Exercises for Section 14.11

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

14.11.1 WP An article in *Rubber Age* (1961, Vol. 89, pp. 453–458) describes an experiment on the manufacture of a product in which two factors were varied. The factors are reaction time (hr) and temperature (°C). These factors are coded as $x_1 = (\text{time} - 12)/8$ and $x_2 = (\text{temperature} - 250)/30$.

The following data were observed where y is the yield (in percent):

Run Number	x_1	x_2	y
1	−1	0	83.8
2	1	0	81.7
3	0	0	82.4
4	0	0	82.9
5	0	−1	84.7
6	0	1	75.9
7	0	0	81.2
8	−1.414	−1.414	81.3
9	−1.414	1.414	83.1
10	1.414	−1.414	85.3
11	1.414	1.414	72.7
12	0	0	82.0

a. Plot the points at which the experimental runs were made.
b. Fit a second-order model to the data. Is the second-order model adequate?
c. Plot the yield response surface. What recommendations would you make about the operating conditions for this process?

14.11.2 WP An article in *Quality Engineering* ["Mean and Variance Modeling with Qualitative Responses: A Case Study" (1998–1999, Vol. 11, pp. 141–148)] studied how three active ingredients of a particular food affect the overall taste of the product. The measure of the overall taste is the overall mean liking score (MLS). The three ingredients are identified by the variables x_1, x_2, and x_3. The data are shown in the following table.

Run	x_1	x_2	x_3	MLS
1	1	1	−1	6.3261
2	1	1	1	6.2444
3	0	0	0	6.5909
4	0	−1	0	6.3409
5	1	−1	1	5.907
6	1	−1	−1	6.488
7	0	0	−1	5.9773
8	0	1	0	6.8605
9	−1	−1	1	6.0455
10	0	0	1	6.3478
11	1	0	0	6.7609
12	−1	−1	−1	5.7727
13	−1	1	−1	6.1805
14	−1	1	1	6.4894
15	−1	0	0	6.8182

a. Fit a second-order response surface model to the data.
b. Construct contour plots and response surface plots for MLS. What are your conclusions?
c. Analyze the residuals from this experiment. Does your analysis indicate any potential problems?
d. This design has only a single center point. Is this a good design in your opinion?

14.11.3 WP Consider the first-order model

$$\hat{y} = 50 + 1.5x_1 - 0.8x_2$$

where $-1 \le x_i \le 1$. Find the direction of steepest ascent.

14.11.4 WP A manufacturer of cutting tools has developed two empirical equations for tool life (y_1) and tool cost (y_2). Both models are functions of tool hardness (x_1) and manufacturing time (x_2). The equations are

$$\hat{y}_1 = 10 + 5x_1 + 2x_2$$

$$\hat{y}_2 = 23 + 3x_1 + 4x_2$$

and both are valid over the range $-1.5 \le x_i \le 1.5$. Suppose that tool life must exceed 12 hours and cost must be below $27.50.

a. Is there a feasible set of operating conditions?
b. Where would you run this process?

14.11.5 WP An article in *Tappi* (1960, Vol. 43, pp. 38–44) describes an experiment that investigated the ash value of paper pulp (a measure of inorganic impurities). Two variables, temperature T in degrees Celsius and time t in hours, were studied, and some of the results are shown in the following table. The coded predictor variables shown are

$$x_1 = \frac{(T - 775)}{115}, \qquad x_2 = \frac{(t - 3)}{1.5}$$

and the response y is (dry ash value in %) × 103.

x_1	x_2	y	x_1	x_2	y
−1	−1	211	0	−1.5	168
1	−1	92	0	1.5	179
−1	1	216	0	0	122
1	1	99	0	0	175
−1.5	0	222	0	0	157
1.5	0	48	0	0	146

a. What type of design has been used in this study? Is the design rotatable?
b. Fit a quadratic model to the data. Is this model satisfactory?
c. If it is important to minimize the ash value, where would you run the process?

14.11.6 WP In their book *Empirical Model Building and Response Surfaces* (John Wiley, 1987), Box and Draper described an experiment with three factors. The data in the following table are a variation of the original experiment from their book. Suppose that these data were collected in a semiconductor manufacturing process.

a. The response y_1 is the average of three readings on resistivity for a single wafer. Fit a quadratic model to this response.

b. The response y_2 is the standard deviation of the three resistivity measurements. Fit a linear model to this response.

c. Where would you recommend that we set x_1, x_2, and x_3 if the objective is to hold mean resistivity at 500 and minimize the standard deviation?

x_1	x_2	x_3	y_1	y_2
−1	−1	−1	24.00	12.49
0	−1	−1	120.33	8.39
1	−1	−1	213.67	42.83
−1	0	−1	86.00	3.46
0	0	−1	136.63	80.41
1	0	−1	340.67	16.17
−1	1	−1	112.33	27.57
0	1	−1	256.33	4.62
1	1	−1	271.67	23.63
−1	−1	0	81.00	0.00
0	−1	0	101.67	17.67
1	−1	0	357.00	32.91
−1	0	0	171.33	15.01
0	0	0	372.00	0.00
1	0	0	501.67	92.50
−1	1	0	264.00	63.50
0	1	0	427.00	88.61
1	1	0	730.67	21.08
−1	−1	1	220.67	133.82
0	−1	1	239.67	23.46
1	−1	1	422.00	18.52
−1	0	1	199.00	29.44
0	0	1	485.33	44.67
1	0	1	673.67	158.21
−1	1	1	176.67	55.51
0	1	1	501.00	138.94
1	1	1	1010.00	142.45

14.11.7 SS Consider the first-order model

$$y = 12 + 1.2x_1 - 2.1x_2 + 1.6x_3 - 0.6x_4$$

where $-1 \le x_i \le 1$.

a. Find the direction of steepest ascent.

b. Assume that the current design is centered at the point (0, 0, 0, 0). Determine the point that is three units from the current center point in the direction of steepest ascent.

14.11.8 Suppose that a response y_1 is a function of two inputs x_1 and x_2 with $y_1 = 2x_2^2 - 4x_1^2 - x_1x_2 + 4$.

a. Draw the contours of this response function.

b. Consider another response $y_2 = (x_1 - 2)^2 + (x_2 - 3)^2$.

c. Add the contours for y_2 and discuss how feasible it is to minimize both y_1 and y_2 with values for x_1 and x_2.

14.11.9 Two responses y_1 and y_2 are related to two inputs x_1 and x_2 by the models $y_1 = 5 + (x_1 - 2)^2 + (x_2 - 3)^2$ and $y_2 = x_2 - x_1 + 3$. Suppose that the objectives are $y_1 \le 9$ and $y_2 \ge 6$.

a. Is there a feasible set of operating conditions for x_1 and x_2? If so, plot the feasible region in the space of x_1 and x_2.

b. Determine the point(s) (x_1, x_2) that yields $y_2 \ge 6$ and minimizes y_1.

14.11.10 An article in the *Journal of Materials Processing Technology* (1997, Vol. 67, pp. 55–61) used response surface methodology to generate surface roughness prediction models for turning EN 24T steel (290 BHN). The data are shown in the following table.

The factors and levels for the experiment are shown below.

Levels	**Lowest**	**Low**	**Center**	**High**	**Highest**
Coding	$-\sqrt{2}$	−1	0	1	$\sqrt{2}$
Speed, V (m min^{-1})	28	36	65	117	150
Feed, f (mm rev^{-1})	0.12	0.15	0.25	0.40	0.50
Depth of cut, d (mm)	0.42	0.50	0.75	1.125	1.33

a. Plot the points at which the experimental runs were made.

b. Fit both first-and second-order models to the data. Comment on the adequacies of these models.

c. Plot the roughness response surface for the second-order model and comment.

14.11.11 An article in *Analytical Biochemistry* ["Application of Central Composite Design for DNA Hybridization Onto Magnetic Microparticles," (2009, Vol. 391(1), 2009, pp. 17–23)] considered the effects of probe and target concentration and particle number in immobilization and hybridization on a micro particle-based DNA hybridization assay. Mean fluorescence is the response. Particle concentration was transformed to surface area measurements. Other concentrations were measured in micromoles per liter (μM). Data are shown below.

Trial	Speed (m min^{-1})	Feed (mm rev^{-1})	Depth of cut (mm)	Coding			Surface roughness, (μm)
				x_1	x_2	x_3	
1	36	0.15	0.50	−1	−1	−1	1.8
2	117	0.15	0.50	1	−1	−1	1.233
3	36	0.40	0.50	−1	1	−1	5.3
4	117	0.40	0.50	1	1	−1	5.067
5	36	0.15	1.125	−1	−1	1	2.133
6	117	0.15	1.125	1	−1	1	1.45
7	36	0.40	1.125	−1	1	1	6.233
8	117	0.40	1.125	1	1	1	5.167
9	65	0.25	0.75	0	0	0	2.433
10	65	0.25	0.75	0	0	0	2.3
11	65	0.25	0.75	0	0	0	2.367
12	65	0.25	0.75	0	0	0	2.467
13	28	0.25	0.75	$-\sqrt{2}$	0	0	3.633
14	150	0.25	0.75	$\sqrt{2}$	0	0	2.767
15	65	0.12	0.75	0	$-\sqrt{2}$	0	1.153
16	65	0.50	0.75	0	$\sqrt{2}$	0	6.333
17	65	0.25	0.42	0	0	$-\sqrt{2}$	2.533
18	65	0.25	1.33	0	0	$\sqrt{2}$	3.20
19	28	0.25	0.75	$-\sqrt{2}$	0	0	3.233
20	150	0.25	0.75	$\sqrt{2}$	0	0	2.967
21	65	0.12	0.75	0	$-\sqrt{2}$	0	1.21
22	65	0.50	0.75	0	$\sqrt{2}$	0	6.733
23	65	0.25	0.42	0	0	$-\sqrt{2}$	2.833
24	65	0.25	1.33	0	0	$\sqrt{2}$	3.267

Run	Immobilization area (cm^2)	Probe area (μM)	Hybridization (cm^2)	Target (μM)	Mean Fluorescence
1	0.35	0.025	0.35	0.025	4.7
2	7	0.025	0.35	0.025	4.7
3	0.35	2.5	0.35	0.025	28.0
4	7	2.5	0.35	0.025	81.2
5	0.35	0.025	3.5	0.025	5.7

6	7	0.025	3.5	0.025	3.8
7	0.35	2.5	3.5	0.025	12.2
8	7	2.5	3.5	0.025	19.5
9	0.35	0.025	0.35	5	4.4
10	7	0.025	0.35	5	2.6
11	0.35	2.5	0.35	5	83.7
12	7	2.5	0.35	5	84.7
13	0.35	0.025	3.5	5	6.8
14	7	0.025	3.5	5	2.4
15	0.35	2.5	3.5	5	76
16	7	2.5	3.5	5	77.9
17	0.35	5	2	2.5	42.6
18	7	5	2	2.5	52.3
19	3.5	0.025	2	2.5	2.6
20	3.5	2.5	2	2.5	72.8
21	3.5	5	0.35	2.5	47.7
22	3.5	5	3.5	2.5	54.4
23	3.5	5	2	0.025	30.8
24	3.5	5	2	5	64.8
25	3.5	5	2	2.5	51.6
26	3.5	5	2	2.5	52.6
27	3.5	5	2	2.5	56.1

a. What type of design is used?
b. Fit a second-order response surface model to the data.
c. Does a residual analysis indicate any problems?

14.11.12 An article in *Applied Biochemistry and Biotechnology* ["A Statistical Approach for Optimization of Polyhydroxybutyrate Production by Bacillus sphaericus NCIM 5149 under Submerged Fermentation Using Central Composite Design" (2010, Vol. 162(4), pp. 996–1007)] described an experiment to optimize the production of polyhydroxybutyrate (PHB). Inoculum age, pH, and substrate were selected as factors, and a central composite design was conducted. Data follow.

Run	Inoculum age (h)	pH	Substrate (g/L)	PHB (g/L)
1	12	4	1	0.84
2	24	8	1	0.55
3	18	6	2.5	1.96
4	28	6	2.5	1.2
5	12	4	4	0.783
6	18	6	2.5	1.66
7	18	6	2.5	2.22
8	18	6	5	0.8

9	12	8	4	0.48
10	18	6	2.5	1.97
11	18	6	2.5	2.2
12	18	6	2.5	2.25
13	18	2	2.5	0.2
14	18	6	0	0.22
15	12	8	1	0.37
16	24	8	4	0.66
17	24	4	1	0.28
18	24	4	4	0.88
19	18	9	2.5	0.3
20	7	6	2.5	0.42

a. Plot the points at which the experimental runs were made. [*Hint*: Code each variable first.] What type of design is used?

b. Fit a second-order response surface model to the data.

c. Does a residual analysis indicate any problems?

d. Construct a contour plot and response surface for PHB amount in terms of two factors.

e. Can you recommend values for inoculum age, pH and substrate to maximize production?

Supplemental Exercises for Chapter 14

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

14.S6 WP Heat-treating metal parts is a widely used manufacturing process. An article in the *Journal of Metals* ["Optimizing Heat Treatment with Factorial Design" (1989, Vol. 41(3), pp. 52–53)] described an experiment to investigate flatness distortion from heat-treating for three types of gears and two heat-treating times. The data follow:

	Time (minutes)	
Gear Type	**90**	**120**
20-tooth	0.0265	0.0560
	0.0340	0.0650
24-tooth	0.0430	0.0720
	0.0510	0.0880
28-tooth	0.0405	0.0620
	0.0575	0.0825

a. Is there any evidence that flatness distortion is different for the different gear types? Is there any indication that heat-treating time affects the flatness distortion? Do these factors interact? Use $\alpha = 0.05$.

b. Construct graphs of the factor effects that aid in drawing conclusions from this experiment.

c. Analyze the residuals from this experiment. Comment on the validity of the underlying assumptions.

14.S7 WP An article in *Process Engineering* (1992, No. 71, pp. 46–47) presented a two-factor factorial experiment to investigate the effect of pH and catalyst concentration on product viscosity (cSt). The data are as follows:

	Catalyst Concentration	
pH	**2.5**	**2.7**
5.6	192, 199, 189, 198	178, 186, 179, 188
5.9	185, 193, 185, 192	197, 196, 204, 204

a. Test for main effects and interactions using $\alpha = 0.05$. What are your conclusions?

b. Graph the interaction and discuss the information provided by this plot.

c. Analyze the residuals from this experiment.

14.S8 WP An article in the *IEEE Transactions on Components, Hybrids, and Manufacturing Technology* (1992, Vol. 15) described an experiment for aligning optical chips onto circuit boards. The method involves placing solder bumps onto the bottom of the chip. The experiment used three solder bump sizes and three alignment methods. The response variable is alignment accuracy (in micrometers). The data are as follows:

	Alignment Method		
Solder Bump Size (diameter in mm)	**1**	**2**	**3**
75	4.60	1.55	1.05
	4.53	1.45	1.00
130	2.33	1.72	0.82
	2.44	1.76	0.95
260	4.95	2.73	2.36
	4.55	2.60	2.46

a. Is there any indication that either solder bump size or alignment method affects the alignment accuracy? Is there any evidence of interaction between these factors? Use $\alpha = 0.05$.

b. What recommendations would you make about this process?

c. Analyze the residuals from this experiment. Comment on model adequacy.

14.S9 An article in the *Textile Research Institute Journal* (1984, Vol. 54, pp. 171–179) reported the results of an experiment that studied the effects of treating fabric with selected inorganic salts on the flammability of the material. Two application levels of each salt were used, and a vertical burn test was used on each sample. (This finds the temperature at which each sample ignites.) The burn test data follow.

	Salt					
Level	**Untreated**	**$MgCl_2$**	**NaCl**	**$CaCO_3$**	**$CaCl_2$**	**Na_2CO_3**
1	812	752	739	733	725	751
	827	728	731	728	727	761
	876	764	726	720	719	755
2	945	794	741	786	756	910
	881	760	744	771	781	854
	919	757	727	779	814	848

a. Test for differences between salts, application levels, and interactions. Use $\alpha = 0.01$.

b. Draw a graph of the interaction between salt and application level. What conclusions can you draw from this graph?

c. Analyze the residuals from this experiment.

14.S10 WP An article in the *Journal of Coatings Technology* (1988, Vol. 60, pp. 27–32) described a 2^4 factorial design used for studying a silver automobile basecoat. The response variable is distinctness of image (DOI). The variables used in the experiment are

A = Percentage of polyester by weight of polyester/melamine (low value = 50%, high value = 70%)

B = Percentage of cellulose acetate butyrate carboxylate (low value = 15%, high value = 30%)

C = Percentage of aluminum stearate (low value = 1%, high value = 3%)

D = Percentage of acid catalyst (low value = 0.25%, high value = 0.50%)

The responses are (1) = 63.8, a = 77.6, b = 68.8, ab = 76.5, c = 72.5, ac = 77.2, bc = 77.7, abc = 84.5, d = 60.6, ad = 64.9, bd = 72.7, abd = 73.3, cd = 68.0, acd = 76.3, bcd = 76.0, and $abcd$ = 75.9.

a. Estimate the factor effects.

b. From a normal probability plot of the effects, identify a tentative model for the data from this experiment.

c. Using the apparently negligible factors as an estimate of error, test for significance of the factors identified in part (b). Use $\alpha = 0.05$.

d. What model would you use to describe the process based on this experiment? Interpret the model.

e. Analyze the residuals from the model in part (d) and comment on your findings.

14.S11 WP SS An article in *Solid State Technology* (1984, Vol. 29, pp. 281–284) described the use of factorial experiments in photolithography, an important step in the process of manufacturing integrated circuits. The variables in this experiment (all at two levels) are prebake temperature (A), prebake time (B), and exposure energy (C), and the response variable is delta line width, the difference between the line on the mask and the printed line on the device. The data are as follows: (1) = −2.30, a = −9.87, b = −18.20, ab = −30.20, c = −23.80, ac = −4.30, bc = −3.80, and abc = −14.70.

a. Estimate the factor effects.

b. Use a normal probability plot of the effect estimates to identity factors that may be important.

c. What model would you recommend for predicting the delta line width response based on the results of this experiment?

d. Analyze the residuals from this experiment, and comment on model adequacy.

14.S12 Construct a 2^{4-1}_{IV} design for the problem in Exercise 14.S10. Select the data for the eight runs that would have been required for this design. Analyze these runs and compare your conclusions to those obtained in Exercise 14.S10 for the full factorial.

14.S13 Construct a 2^{5-2}_{III} design in eight runs. What are the alias relationships in this design?

14.S14 Construct a 2^{8-4}_{IV} design in 16 runs. What are the alias relationships in this design?

14.S15 An article in *Rubber Chemistry and Technology* (1974, Vol. 47, pp. 825–836) described an experiment to study the effect of several variables on the Mooney viscosity of rubber, including silica filler (parts per hundred) and oil filler (parts per hundred). Data typical of that reported in this experiment are reported in the following table where

$$x_1 = \frac{\text{silica} - 60}{15}, \qquad x_2 = \frac{\text{oil} - 21}{15}$$

a. What type of experimental design has been used?

b. Analyze the data and draw appropriate conclusions.

Coded levels		
x_1	x_2	y
−1	−1	13.71
1	−1	14.15
−1	1	12.87
1	1	13.53
−1	−1	13.90
1	−1	14.88
−1	1	12.25
−1	1	13.35

14.S16 WP An article in *Plant Disease* ["Effect of Nitrogen and Potassium Fertilizer Rates on Severity of Xanthomonas Blight of *Syngonium Podophyllum*" (1989, Vol. 73(12), pp. 972–975)] showed the effect of the variable nitrogen and potassium rates on the growth of "White Butterfly" and the mean percentage of disease. Data representative of that collected in this experiment is provided in the following table.

	Potassium (mg/pot/wk)					
Nitrogen (mg/pot/wk)	**30**		**90**		**120**	
50	61.0	61.3	45.5	42.5	59.5	58.2
150	54.5	55.9	53.5	51.9	34.0	35.9
250	42.7	40.4	36.5	37.4	32.5	33.8

a. State the appropriate hypotheses.
b. Use the analysis of variance to test these hypotheses with $\alpha = 0.05$.
c. Graphically analyze the residuals from this experiment.
d. Estimate the appropriate variance components.

14.S17 SS An article in the *Journal of Applied Electrochemistry* (May 2008, Vol. 38(5), pp. 583–590) presented a 2^{7-3} fractional factorial design to perform optimization of polybenzimidazole-based membrane electrode assemblies for H_2/O_2 fuel cells. The design and data are shown in the following table.

Runs	***A***	***B***	***C***	***D***	***E***	***F***	***G***	**Current Density (CD mA cm²)**
1	−1	−1	−1	−1	−1	−1	−1	160
2	+1	−1	−1	−1	+1	+1	+1	20
3	−1	+1	−1	−1	+1	+1	−1	80
4	+1	+1	−1	−1	−1	−1	+1	317
5	−1	−1	+1	−1	+1	−1	+1	19
6	+1	−1	+1	−1	−1	+1	−1	4
7	−1	+1	+1	−1	−1	+1	+1	20
8	+1	+1	+1	−1	+1	−1	−1	88
9	−1	−1	−1	+1	−1	+1	+1	1100
10	+1	−1	−1	+1	+1	−1	−1	12
11	−1	+1	−1	+1	+1	−1	+1	552
12	+1	+1	−1	+1	−1	+1	−1	880
13	−1	−1	+1	+1	+1	+1	−1	16
14	+1	−1	+1	+1	−1	−1	+1	20
15	−1	+1	+1	+1	−1	−1	−1	8
16	+1	+1	+1	+1	+1	+1	+1	15

The factors and levels are shown in the following table.

	Factor	**−1**	**+1**
A	Amount of binder in the catalyst layer	0.2 mg cm²	1 mg cm²
B	Electrocatalyst loading	0.1 mg cm²	1 mg cm²
C	Amount of carbon in the gas diffusion layer	2 mg cm²	4.5 mg cm²
D	Hot compaction time	1 min	10 min
E	Compaction temperature	100°C	150°C
F	Hot compaction load	0.04 ton cm²	0.2 ton cm²
G	Amount of PTFE in the gas diffusion layer	0.1 mg cm²	1 mg cm²

a. Write down the alias relationships.
b. Estimate the main effects.
c. Prepare a normal probability plot for the effects and interpret the results.
d. Calculate the sum of squares for the alias set that contains the ABG interaction from the corresponding effect estimate.

14.S18 Consider the following results from a two-factor experiment with two levels for factor *A* and three levels for factor *B*. Each treatment has three replicates.

A	***B***	**Mean**	**StDev**
1	1	21.33333	6.027714
1	2	20	7.549834
1	3	32.66667	3.511885
2	1	31	6.244998
2	2	33	6.557439
2	3	23	10

a. Calculate the sum of squares for each factor and the interaction.
b. Calculate the sum of squares total and error.
c. Complete an ANOVA table with *F*-statistics.

14.S19 The rework time required for a machine was found to depend on the speed at which the machine was run (*A*), the lubricant used while working (*B*), and the hardness of the metal used in the machine (*C*). Two levels of each factor were chosen and a single replicate of a 2^3 experiment was run. The rework time data (in hours) are shown in the following table.

Treatment Combination	**Time (in hours)**
(1)	27
a	34
b	38
ab	59
c	44
ac	40
bc	63
abc	37

a. These treatments cannot all be run under the same conditions. Set up a design to run these observations in two blocks of four observations each, with *ABC* confounded with blocks.

b. Analyze the data.

14.S20 An article in *Process Biochemistry* (1996, Vol. 31(8), pp. 773–785) presented a 2^{7-3} fractional factorial to perform optimization of manganese dioxide bioleaching media. The data are shown in the following table.

Runs	*A*	*B*	*C*	*D*	*E*	*F*	*G*	Manganese Extraction Yield (%)
1	−1	−1	−1	−1	−1	−1	−1	99.0
2	1	−1	−1	−1	1	−1	1	97.4
3	−1	1	−1	−1	1	1	1	97.7
4	1	1	−1	−1	−1	1	−1	90.0
5	−1	−1	1	−1	1	1	−1	100.0
6	1	−1	1	−1	−1	1	1	98.0
7	−1	1	1	−1	−1	−1	1	90.0
8	1	1	1	−1	1	−1	−1	93.5
9	−1	−1	−1	1	−1	1	1	100.0
10	1	−1	−1	1	1	1	−1	98.6
11	−1	1	−1	1	1	−1	−1	97.1
12	1	1	−1	1	−1	−1	1	92.4
13	−1	−1	1	1	1	−1	1	93.0
14	1	−1	1	1	−1	−1	−1	95.0
15	−1	1	1	1	−1	1	−1	97.0
16	1	1	1	1	1	1	1	98.0

The factors and levels are shown in the following table.

	Factor	−1	+1
A	Mineral concentration (%)	10	20
B	Molasses (g/liter)	100	200
C	NH_4NO_3 (g/liter)	1.25	2.50
D	KH_2PO_4 (g/liter)	0.75	1.50
E	$MgSO_4$ (g/liter)	0.5	1.00
F	Yeast extract (g/liter)	0.20	0.50
G	$NaHCO_3$ (g/liter)	2.00	4.00

a. Write down the complete defining relation and the aliases from the design.

b. Estimate the main effects.

c. Plot the effect estimates on normal probability paper and interpret the results.

d. Conduct a residual analysis.

14.S21 Consider the following ANOVA table from a two-factor factorial experiment.

Two-way ANOVA: y Versus A, B

Source	DF	SS	MS	F	P
A	3	1213770	404590	?	0.341
B	2	?	17335441	58.30	0.000
Error	?	1784195	?		
Total	11	37668847			

a. How many levels of each factor were used in the experiment?

b. How many replicates were used?

c. What assumption is made in order to obtain an estimate of error?

d. Calculate the missing entries (denoted with "?") in the ANOVA table.

Chapter 15 Exercises

Exercises for Section 15.3

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

15.3.1 WP Control charts are to be constructed for samples of size $n = 4$, and $\bar{x}$ and s are computed for each of 20 preliminary samples as follows:

$$\sum_{i=1}^{20} \bar{x}_i = 4460 \qquad \sum_{i=1}^{20} s_i = 271.6$$

a. Calculate trial control limits for $\bar{X}$ and S charts.

b. Assuming the process is in control, estimate the process mean and standard deviation.

15.3.2 WP Twenty-five samples of size 5 are drawn from a process at one-hour intervals, and the following data are obtained:

$$\sum_{i=1}^{25} \bar{x}_i = 362.75 \qquad \sum_{i=1}^{25} r_i = 8.60 \qquad \sum_{i=1}^{25} s_i = 3.64$$

a. Calculate trial control limits for $\overline{X}$ and R charts.
b. Repeat part (a) for $\overline{X}$ and S charts.

15.3.3 WP The level of cholesterol (in mg/dL) is an important index for human health. The sample size is $n = 5$. The following summary statistics are obtained from cholesterol measurements:

$$\sum_{i=1}^{30} \bar{x}_i = 140.03, \quad \sum_{i=1}^{30} r_i = 13.63, \quad \sum_{i=1}^{30} s_i = 5.10$$

a. Find trial control limits for $\overline{X}$ and R charts.
b. Repeat part (a) for $\overline{X}$ and S charts.

15.3.4 An $\overline{X}$ control chart with 3-sigma control limits has $UCL = 48.75$ and $LCL = 42.71$. Suppose that the process standard deviation is $\sigma = 2.25$. What subgroup size was used for the chart?

15.3.5 WP SS The pull strength of a wire-bonded lead for an integrated circuit is monitored. The following table provides data for 20 samples each of size 3.

Sample Number	x_1	x_2	x_3
1	15.4	15.6	15.3
2	15.4	17.1	15.2
3	16.1	16.1	13.5
4	13.5	12.5	10.2
5	18.3	16.1	17.0
6	19.2	17.2	19.4
7	14.1	12.4	11.7
8	15.6	13.3	13.6
9	13.9	14.9	15.5
10	18.7	21.2	20.1
11	15.3	13.1	13.7
12	16.6	18.0	18.0
13	17.0	15.2	18.1
14	16.3	16.5	17.7
15	8.4	7.7	8.4
16	11.1	13.8	11.9
17	16.5	17.1	18.5
18	18.0	14.1	15.9
19	17.8	17.3	12.0
20	11.5	10.8	11.2

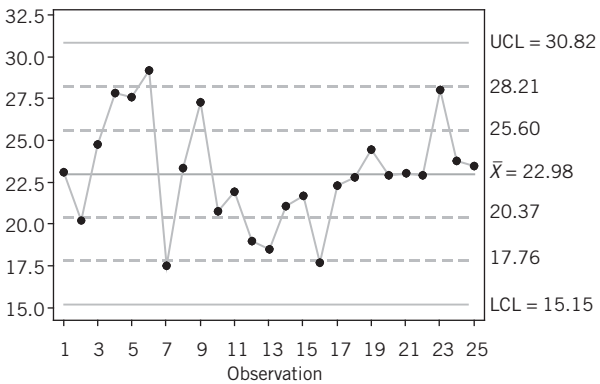
a. Use all the data to determine trial control limits for $\overline{X}$ and R charts, construct the control limits, and plot the data.
b. Use the control limits from part (a) to identify out-of-control points. If necessary, revise your control limits assuming that any samples that plot outside of the control limits can be eliminated.
c. Repeat parts (a) and (b) for $\overline{X}$ and S charts.

15.3.6 WP VS The copper content of a plating bath is measured three times per day, and the results are reported in ppm. The $\bar{x}$ and r values for 25 days are shown in the following table:

Day	$\bar{x}$	r	Day	$\bar{x}$	r
1	5.45	1.21	14	7.01	1.45
2	5.39	0.95	15	5.83	1.37
3	6.85	1.43	16	6.35	1.04
4	6.74	1.29	17	6.05	0.83
5	5.83	1.35	18	7.11	1.35
6	7.22	0.88	19	7.32	1.09
7	6.39	0.92	20	5.90	1.22
8	6.50	1.13	21	5.50	0.98
9	7.15	1.25	22	6.32	1.21
10	5.92	1.05	23	6.55	0.76
11	6.45	0.98	24	5.90	1.20
12	5.38	1.36	25	5.95	1.19
13	6.03	0.83			

a. Using all the data, find trial control limits for $\overline{X}$ and R charts, construct the chart, and plot the data. Is the process in statistical control?
b. If necessary, revise the control limits computed in part (a), assuming that any samples that plot outside the control limits can be eliminated.

15.3.7 Apply the Western Electric rules to the following control chart. The warning limits are shown as dotted lines. Describe any rule violations.



15.3.8 The following data were considered in *Quality Engineering* ["An SPC Case Study on Stabilizing Syringe Lengths" (1999–2000, Vol. 12(1), pp. 65–71)]. The syringe length is measured during a pharmaceutical manufacturing process. The following table provides data (in inches) for 20 samples each of size 5.

Sample	x_1	x_2	x_3	x_4	x_5
1	4.960	4.946	4.950	4.956	4.958
2	4.958	4.927	4.935	4.940	4.950
3	4.971	4.929	4.965	4.952	4.938
4	4.940	4.982	4.970	4.953	4.960
5	4.964	4.950	4.953	4.962	4.956
6	4.969	4.951	4.955	4.966	4.954
7	4.960	4.944	4.957	4.948	4.951
8	4.969	4.949	4.963	4.952	4.962
9	4.984	4.928	4.960	4.943	4.955
10	4.970	4.934	4.961	4.940	4.965
11	4.975	4.959	4.962	4.971	4.968
12	4.945	4.977	4.950	4.969	4.954
13	4.976	4.964	4.970	4.968	4.972
14	4.970	4.954	4.964	4.959	4.968
15	4.982	4.962	4.968	4.975	4.963
16	4.961	4.943	4.950	4.949	4.957
17	4.980	4.970	4.975	4.978	4.977
18	4.975	4.968	4.971	4.969	4.972
19	4.977	4.966	4.969	4.973	4.970
20	4.975	4.967	4.969	4.972	4.972

a. Using all the data, find trial control limits for $\overline{X}$ and R charts, construct the chart, and plot the data. Is this process in statistical control?

b. Use the trial control limits from part (a) to identify out-of-control points. If necessary, revise your control limits assuming that any samples that plot outside the control limits can be eliminated.

c. Repeat parts (a) and (b) for $\overline{X}$ and S charts.

15.3.9 SS Consider the data in Exercise 15.3.5. Calculate the sample standard deviation of all 60 measurements and compare this result to the estimate of σ obtained from your revised $\overline{X}$ and R charts. Explain any differences.

15.3.10 Web traffic can be measured to help highlight security problems or indicate a potential lack of bandwidth. Data on Web traffic (in thousand hits) from http://en.wikipedia.org/wiki/Web_traffic are given in the following table for 25 samples each of size 4.

Sample	x_1	x_2	x_3	x_4
1	163.95	164.54	163.87	165.10
2	163.30	162.85	163.18	165.10
3	163.13	165.14	162.80	163.81
4	164.08	163.43	164.03	163.77
5	165.44	163.63	163.95	164.78
6	163.83	164.14	165.22	164.91
7	162.94	163.64	162.30	163.78
8	164.97	163.68	164.73	162.32
9	165.04	164.06	164.40	163.69
10	164.74	163.74	165.10	164.32
11	164.72	165.75	163.07	163.84
12	164.25	162.72	163.25	164.14
13	164.71	162.63	165.07	162.59
14	166.61	167.07	167.41	166.10
15	165.23	163.40	164.94	163.74
16	164.27	163.42	164.73	164.88
17	163.59	164.84	164.45	164.12
18	164.90	164.20	164.32	163.98
19	163.98	163.53	163.34	163.82
20	164.08	164.33	162.38	164.08
21	165.71	162.63	164.42	165.27
22	164.03	163.36	164.55	165.77
23	160.52	161.68	161.18	161.33
24	164.22	164.27	164.35	165.12
25	163.93	163.96	165.05	164.52

a. Use all the data to determine trial control limits for $\overline{X}$ and R charts, construct the chart, and plot the data.

b. Use the trial control limits from part (a) to identify out-of-control points. If necessary, revise your control limits, assuming that any samples that plot outside the control limits can be eliminated.

15.3.11 An $\overline{X}$ control chart with 3-sigma control limits and subgroup size $n = 4$ has control limits $UCL = 48.75$ and $LCL = 40.55$.

a. Estimate the process standard deviation.

b. Does the response to part (a) depend on whether $\overline{r}$ or $\overline{s}$ was used to construct the $\overline{X}$ control chart?

15.3.12 An article in *Quality & Safety in Health Care* ["Statistical Process Control as a Tool for Research and Healthcare Improvement" (2003, Vol. 12, pp. 458–464)] considered a

number of control charts in healthcare. The following approximate data were used to construct $\overline{X} - S$ charts for the turnaround time (TAT) for complete blood counts (in minutes). The subgroup size is $n = 3$ per shift, and the mean standard deviation is 21. Construct the $\overline{X}$ chart and comment on the control of the process. If necessary, assume that assignable causes can be found, eliminate suspect points, and revise the control limits.

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14
TAT	51	73	28	52	65	49	51	50	25	39	40	30	49	31

Exercises for Section 15.4

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

15.4.1 WP In a semiconductor manufacturing process, CVD metal thickness was measured on 30 wafers obtained over approximately 2 weeks. Data are shown in the following table.

a. Using all the data, compute trial control limits for individual observations and moving-range charts. Construct the chart and plot the data. Determine whether the process is in statistical control. If not, assume that assignable causes can be found to eliminate these samples and revise the control limits.

b. Estimate the process mean and standard deviation for the in-control process.

Wafer	x	Wafer	x	Wafer	x
1	16.8	11	17.1	21	15.2
2	14.9	12	15.9	22	16.7
3	18.3	13	16.4	23	15.2
4	16.5	14	15.8	24	14.7
5	17.1	15	15.4	25	17.9
6	17.4	16	15.4	26	14.8
7	15.9	17	14.3	27	17.0
8	14.4	18	16.1	28	16.2
9	15.0	19	15.8	29	15.6
10	15.7	20	15.9	30	16.3

15.4.2 WP The viscosity of a chemical intermediate is measured every hour. Twenty samples each of size $n = 1$ are in the following table.

Sample	Viscosity	Sample	Viscosity	Sample	Viscosity
1	495	8	504	15	497
2	491	9	542	16	499
3	501	10	508	17	468
4	501	11	493	18	486
5	512	12	507	19	511
6	540	13	503	20	487
7	492	14	475		

a. Using all the data, compute trial control limits for individual observations and moving-range charts. Determine whether the process is in statistical control. If not, assume that assignable causes can be found to eliminate these samples and revise the control limits.

b. Estimate the process mean and standard deviation for the in-control process.

15.4.3 WP VS The following table of data was analyzed in *Quality Engineering* [1991–1992, Vol. 4(1)]. The average particle size of raw material was obtained from 25 successive samples.

Observation	Size	Observation	Size	Observation	Size
1	96.1	10	96.8	19	87.4
2	94.4	11	100.9	20	96.1
3	116.2	12	117.7	21	97.1
4	98.8	13	115.6	22	95.7
5	95.0	14	100.5	23	94.2
6	120.3	15	103.1	24	102.4
7	104.8	16	93.1	25	131.9
8	88.4	17	93.7		
9	106.8	18	72.4		

a. Using all the data, compute trial control limits for individual observations and moving-range charts. Construct the chart and plot the data. Determine whether the process is in statistical control. If not, assume that assignable causes can be found to eliminate these samples and revise the control limits.

b. Estimate the process mean and standard deviation for the in-control process.

15.4.4 Pulsed laser deposition technique is a thin film deposition technique with a high-powered laser beam. Twenty-five films were deposited through this technique. The thicknesses of the films obtained are shown in the following table.

Film	Thickness (nm)	Film	Thickness (nm)	Film	Thickness (nm)
1	28	10	35	19	56
2	45	11	47	20	49
3	34	12	50	21	21
4	29	13	32	22	62
5	37	14	40	23	34
6	52	15	46	24	31
7	29	16	59	25	98
8	51	17	20		
9	23	18	33		

a. Using all the data, compute trial control limits for individual observations and moving-range charts. Determine whether the process is in statistical control. If not, assume that assignable causes can be found to eliminate these samples and revise the control limits.

b. Estimate the process mean and standard deviation for the in-control process.

15.4.5 SS An article in the *Journal of the Operational Research Society* ["A Quality Control Approach for Monitoring Inventory Stock Levels" (1993, Vol. 44, pp. 1115–1127)] reported on a control chart to monitor the accuracy of an inventory management system. Inventory accuracy at time t, $AC(t)$, is defined as the difference between the recorded and actual inventory (in absolute value) divided by the recorded inventory. Consequently, $AC(t)$ ranges between 0 and 1, with lower values better. Extracted data are shown in the following table.

t	1	2	3	4	5	6	7	8
$AC(t)$	0.190	0.050	0.095	0.055	0.090	0.200	0.030	0.105
t	9	10	11	12	13	14	15	16
$AC(t)$	0.115	0.103	0.121	0.089	0.180	0.122	0.098	0.173
t	17	18	19	20	21	22		
$AC(t)$	0.298	0.075	0.083	0.115	0.147	0.079		

a. Calculate individuals and moving-range charts for these data.

b. Comment on the control of the process. If necessary, assume that assignable causes can be found, eliminate suspect points, and revise the control limits.

15.4.6 An article in *Quality & Safety in Health Care* ["Statistical Process Control as a Tool for Research and Healthcare Improvement" (2003, Vol. 12, pp. 458–464)] considered a number of control charts in healthcare. An X chart was constructed for the amount of infectious waste discarded each day (in pounds). The article mentions that improperly classified infectious waste (actually not hazardous) adds substantial costs to hospitals each year. The following tables show approximate data for the average daily waste per month before and after process changes, respectively. The process change included an education campaign to provide an operational definition for infectious waste.

Before Process Change

Month	1	2	3	4	5	6	7	8	9
Waste	6.9	6.8	6.9	6.7	6.9	7.5	7	7.4	7
Month	13	14	15	16	17	18	19	20	21
Waste	7.5	7.4	6.5	6.9	7.0	7.2	7.8	6.3	6.7

After Process Change

Month	1	2	3	4	5	6	7	8	9	10
Waste	5.0	4.8	4.4	4.3	4.6	4.3	4.5	3.5	4.0	4.1
Month	11	12	13	14	15	16	17	18	19	20
Waste	3.8	5.0	4.6	4.0	5.0	4.9	4.9	5.0	6.0	4.5
Month	21	22	23	24	25	26	27	28	29	30
Waste	4.0	5.0	4.5	4.6	4.6	3.8	5.3	4.5	4.4	3.8

a. Handle the data before and after the process change separately and construct individuals and moving-range charts for each set of data. Assume that assignable causes can be found and eliminate suspect observations. If necessary, revise the control limits.

b. Comment on the control of each chart and differences between the charts. Was the process change effective?

Exercises for Section 15.5

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

15.5.1 WP Suppose that a quality characteristic is normally distributed with specifications from 20 to 32 units.

a. What value is needed for σ to achieve a PCR of 1.5?

b. What value for the process mean minimizes the fraction defective? Does this choice for the mean depend on the value of σ?

15.5.2 WP Suppose that a quality characteristic is normally distributed with specifications at 100 ± 20. The process standard deviation is 6.

a. Suppose that the process mean is 100. What are the natural tolerance limits? What is the fraction defective? Calculate PCR and PCR_k and interpret these ratios.

b. Suppose that the process mean is 106. What are the natural tolerance limits? What is the fraction defective? Calculate PCR and PCR_k and interpret these ratios.

15.5.3 WP SS A normally distributed process uses 66.7% of the specification band. It is centered at the nominal dimension, located halfway between the upper and lower specification limits.

a. Estimate PCR and PCR_k. Interpret these ratios.

b. What fallout level (fraction defective) is produced?

15.5.4 Reconsider Exercise 15.3.2 in which the specification limits are 14.50 ± 0.50.

a. What conclusions can you draw about the ability of the process to operate within these limits? Estimate the percentage of defective items that is produced.

b. Estimate PCR and PCR_k. Interpret these ratios.

15.5.5 WP SS Reconsider Exercise 15.3.1. Suppose that the variable is normally distributed with specifications at 220 ± 50. What is the proportion out of specifications? Estimate and interpret PCR and PCR_k.

15.5.6 Reconsider the copper-content measurements in Exercise 15.3.6. Given that the specifications are at 6.0 ± 1.0, estimate PCR and PCR_k and interpret these ratios.

15.5.7 VS Suppose that a quality characteristic is normally distributed with specifications at 120 ± 20. The process standard deviation is 6.5.

a. Suppose that the process mean is 120. What are the natural tolerance limits? What is the fraction defective? Calculate PCR and PCR_k and interpret these ratios.

b. Suppose that the process mean shifts off-center by 1.5 standard deviations toward the upper specification limit. Recalculate the quantities in part (a).

c. Compare the results in parts (a) and (b) and comment on any differences.

15.5.8 WP Reconsider the viscosity measurements in Exercise 15.4.2. The specifications are 500 ± 25. Calculate estimates of the process capability ratios PCR and PCR_k for this process and provide an interpretation.

15.5.9 A process mean is centered between the specification limits and $PCR = 1.33$. Assume that the process mean increases by 1.5σ.

a. Calculate PCR and PCR_k for the shifted process.

b. Calculate the estimated fallout from the shifted process and compare your result to those in Table 15.4. Assume a normal distribution for the measurement.

15.5.10 The PCR for a measurement is 1.5 and the control limits for an $\overline{X}$ chart with $n = 4$ are 24.6 and 32.6.

a. Estimate the process standard deviation σ.

b. Assume that the specification limits are centered around the process mean. Calculate the specification limits.

Exercises for Section 15.6

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

15.6.1 WP VS An early example of SPC was described in *Industrial Quality Control* ["The Introduction of Quality Control at Colonial Radio Corporation" (1944, Vol. 1(1), pp. 4–9)]. The following are the fractions defective of shaft and washer assemblies during the month of April in samples of $n = 1500$ each:

Sample	Fraction Defective	Sample	Fraction Defective	Sample	Fraction Defective
1	0.11	8	0.03	15	0.04
2	0.06	9	0.02	16	0.04
3	0.1	10	0.03	17	0.04
4	0.11	11	0.03	18	0.03
5	0.14	12	0.03	19	0.06
6	0.11	13	0.04	20	0.06
7	0.14	14	0.07		

a. Set up a P chart for this process. Is this process in statistical control?

b. Suppose that instead of $n = 1500$, $n = 100$. Use the data given to set up a P chart for this process. Revise the control limits if necessary.

c. Compare your control limits for the P charts in parts (a) and (b). Explain why they differ. Also, explain why your assessment about statistical control differs for the two sizes of n.

15.6.2 WP The following represent the number of defects per 1000 feet in copper wire: 1, 1, 3, 7, 8, 10, 5, 13, 0, 19, 24, 6, 9, 11, 15, 8, 3, 6, 7, 4, 9, 20, 11, 7, 18, 10, 6, 4, 0, 9, 7, 3, 1, 8, 12. Do the data come from a controlled process?

15.6.3 WP SS The following represent the number of solder defects observed on 24 samples of five printed circuit boards: 7, 6, 8, 10, 24, 6, 5, 4, 8, 11, 15, 8, 4, 16, 11, 12, 8, 6, 5, 9, 7, 14, 8, 21.

a. Using all the data, compute trial control limits for a U control chart, construct the chart, and plot the data.

b. Can we conclude that the process is in control using a U chart? If not, assume that assignable causes can be found, and list points and revise the control limits.

15.6.4 The following data are the number of spelling errors detected for every 1000 words on a news Web site over 20 weeks.

Week	No. of Spelling Errors	Week	No. of Spelling Errors
1	3	11	1
2	6	12	6
3	0	13	9
4	5	14	8
5	9	15	6
6	5	16	4
7	2	17	13
8	2	18	3
9	3	19	0
10	2	20	7

a. What control chart is most appropriate for these data?
b. Using all the data, compute trial control limits for the chart in part (a), construct the chart, and plot the data.
c. Determine whether the process is in statistical control. If not, assume that assignable causes can be found and out-of-control points eliminated. Revise the control limits.

15.6.5 A article in *Epilepsy Research* ["Statistical Process Control (SPC): A Simple Objective Method for Monitoring Seizure Frequency and Evaluating Effectiveness of Drug Interventions in Refractory Childhood Epilepsy" (2010, Vol. 91, pp. 205–213)] used control charts to monitor weekly seizure changes in patients with refractory childhood epilepsy. The following table shows representative data of weekly observations of seizure frequency (SF).

Week	1	2	3	4	5	6	7	8	9	10
SF	13	10	17	10	18	14	10	12	16	13
Week	11	12	13	14	15	16	17	18	19	20
SF	14	11	8	11	10	3	2	13	15	21
Week	21	22	23	24	25					
SF	15	12	14	18	12					

a. What type of control chart is appropriate for these data? Construct this chart.
b. Comment on the control of the process.
c. If necessary, assume that assignable causes can be found, eliminate suspect points, and revise the control limits.
d. In the publication, the weekly SFs were approximated as normally distributed and an individuals chart was constructed. Construct this chart and compare it to the attribute chart you built in part (a).

15.6.6 A article in *Graefe's Archive for Clinical and Experimental Ophthalmology* ["Statistical Process Control Charts for Ophthalmology" (2011, Vol. 249, pp. 1103–1105)] considered the number of cataract surgery cases by month. The data are shown in the following table.

a. What type of control chart is appropriate for these data? Construct this chart.
b. Comment on the control of the process.
c. If necessary, assume that assignable causes can be found, eliminate suspect points, and revise the control limits.
d. In the publication, the data were approximated as normally distributed and an individuals chart was constructed. Construct this chart and compare it to the attribute chart you built in part (a). Why might an individuals chart be reasonable?

January	February	March	April	May	June	July
61	88	80	68	80	70	60
August	September	October	November	December		
56	72	118	106	60		

Exercises for Section 15.7

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

15.7.1 WP SS VS An X chart uses samples of size 1. The center line is at 100, and the upper and lower 3-sigma limits are at 112 and 88, respectively.
a. What is the process σ?
b. Suppose that the process mean shifts to 96. Find the probability that this shift is detected on the next sample.
c. Find the ARL to detect the shift in part (b).

15.7.2 WP An $\overline{X}$ chart uses samples of size 4. The center line is at 100, and the upper and lower 3-sigma control limits are at 106 and 94, respectively.
a. What is the process σ?
b. Suppose that the process mean shifts to 96. Find the probability that this shift is detected on the next sample.
c. Find the ARL to detect the shift in part (b).

15.7.3 Consider an $\overline{X}$ control chart with $UCL = 0.0635$, $LCL = 0.0624$, and $n = 5$. Suppose that the mean shifts to 0.0625.

a. What is the probability that this shift is detected on the next sample?

b. What is the ARL after the shift?

15.7.4 WP Consider an $\overline{X}$ control chart with $UCL = 14.708$, $LCL = 14.312$, and $n = 5$. Suppose that the mean shifts to 14.6.

a. What is the probability that this shift is detected on the next sample?

b. What is the ARL after the shift?

15.7.5 An $\overline{X}$ chart uses a sample of size 3. The center line is at 200, and the upper and lower 3-sigma control limits are at 212 and 188, respectively. Suppose that the process mean shifts to 195.

a. Find the probability that this shift is detected on the next sample.

b. Find the ARL to detect the shift in part (a).

15.7.6 WP Consider an $\overline{X}$ control chart with $UCL = 17.40$, $LCL = 12.79$, and $n = 3$. Suppose that the mean shifts to 13.

a. What is the probability that this shift is detected on the next sample?

b. What is the ARL after the shift?

15.7.7 SS Consider a P-chart with subgroup size $n = 50$ and center line at 0.12.

a. Calculate the LCL and UCL.

b. Suppose that the true proportion defective changes from 0.12 to 0.18. What is the ARL after the shift? Assume that the sample proportions are approximately normally distributed.

c. Rework part (a) and (b) with $n = 100$ and comment on the difference in ARL. Does the increased sample size change the ARL substantially?

15.7.8 Consider the U chart for printed circuit boards in Example 15.3.3. The center line = 1.6, $UCL = 3.3$, and $n = 5$.

a. Calculate the LCL and UCL.

b. Suppose that the true mean defects per unit shifts from 1.6 to 2.4. What is the ARL after the shift? Assume that the average defects per unit are approximately normally distributed.

c. Rework part (b) if the true mean defects per unit shifts from 1.6 to 2.0 and comment on the difference in ARL.

Exercises for Section 15.8

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

15.8.1 WP The following data were considered in *Quality Engineering* ["Parabolic Control Limits for the Exponentially Weighted Moving Average Control Charts in Quality Engineering" (1992, Vol. 4(4), pp. 487–495)]. In a chemical plant, the data for one of the quality characteristics (viscosity) were obtained for each 12-hour batch completion. The results of 15 consecutive measurements are shown in the following table.

a. Set up a EWMA control chart for this process with $\lambda = 0.2$. Assume that the desired process target is 14.1. Does the process appear to be in control?

Batch	Viscosity	Batch	Viscosity
1	13.3	9	14.6
2	14.5	10	14.1
3	15.3	11	14.3
4	15.3	12	16.1
5	14.3	13	13.1
6	14.8	14	15.5
7	15.2	15	12.6
8	14.9		

b. Suppose that the next five observations are 14.6, 15.3, 15.7, 16.1, and 16.8. Apply the EWMA in part (a) to these new observations. Is there any evidence that the process has shifted out of control?

15.8.2 The purity of a chemical product is measured every 2 hours. The results of 20 consecutive measurements are as follows:

Sample	Purity	Sample	Purity
1	89.11	11	88.55
2	90.59	12	90.43
3	91.03	13	91.04
4	89.46	14	88.17
5	89.78	15	91.23
6	90.05	16	90.92
7	90.63	17	88.86
8	90.75	18	90.87
9	89.65	19	90.73
10	90.15	20	89.78

Use $\sigma = 0.8$ and assume that the process target is 90.

a. Construct an EWMA control chart with $\lambda = 0.2$. Does the process appear to be in control?

b. Construct an EWMA control chart with $\lambda = 0.5$. Compare your results to part (a).

c. Suppose that the next five observations are 90.75, 90.00, 91.15, 90.95, and 90.86. Apply the EWMAs in parts (a) and (b) to these new observations. Is there any evidence that the process has shifted out of control?

15.8.3 Consider an EMWA control chart. The target value for the process is $\mu_0 = 50$ and $\sigma = 2$. Use Table 15.8.

a. If the sample size is $n = 1$, would you prefer an EWMA chart with $\lambda = 0.1$ and $L = 2.81$ or $\lambda = 0.5$ and $L = 3.07$ to detect a shift in the process mean to $\mu = 52$ on average? Why?

b. If the sample size is increased to $n = 4$, which chart in part (a) do you prefer? Why?

c. If an EWMA chart with $\lambda = 0.1$ and $L = 2.81$ is used, what sample size is needed to detect a shift to $\mu = 52$ in approximately three samples on average?

15.8.4 WP The concentration of a chemical product is measured by taking four samples from each batch of material. The average concentration of these measurements for the last 20 batches is shown in the following table:

Batch	Concentration	Batch	Concentration
1	104.5	11	95.4
2	99.9	12	94.5
3	106.7	13	104.5
4	105.2	14	99.7
5	94.8	15	97.7
6	94.6	16	97
7	104.4	17	95.8
8	99.4	18	97.4
9	100.3	19	99
10	100.3	20	102.6

Use $\sigma = 8$ and assume that the desired process target is 100.

a. Construct an EWMA control chart with $\lambda = 0.2$. Does the process appear to be in control?

b. Construct an EWMA control chart with $\lambda = 0.5$. Compare your results to part (a).

c. If the concentration shifted to 104, would you prefer the chart in part (a) or (b)? Explain.

15.8.5 SS Heart rate (in counts/minute) is measured every 30 minutes. The results of 20 consecutive measurements are as follows:

Sample	Heart Rate	Sample	Heart Rate
1	68	11	79
2	71	12	79
3	67	13	78
4	69	14	78
5	71	15	78
6	70	16	79
7	69	17	79
8	67	18	82
9	70	19	82
10	70	20	81

Use $\mu = 70$ and $\sigma = 3$.

a. Construct an EWMA control chart with $\lambda = 0.1$. Use $L = 2.81$. Does the process appear to be in control?

b. Construct an EWMA control chart with $\lambda = 0.5$. Use $L = 3.07$. Compare your results to those in part (a).

c. If the heart rate mean shifts to 76, approximate the ARLs for the charts in parts (a) and (b).

15.8.6 The number of influenza patients (in thousands) visiting hospitals weekly is shown in the following table.

Sample	Number of Patients	Sample	Number of Patients
1	162.27	13	159.989
2	157.47	14	159.09
3	157.065	15	162.699
4	160.45	16	163.89
5	157.993	17	164.247
6	162.27	18	162.7
7	160.652	19	164.859
8	159.09	20	163.65
9	157.442	21	165.99
10	160.78	22	163.22
11	159.138	23	164.338
12	161.08	24	164.83

Use $\mu = 160$ and $\sigma = 2$.

a. Construct an EWMA control chart with $\lambda = 0.1$. Use $L = 2.81$. Does the process appear to be in control?

b. Construct an EWMA control chart with $\lambda = 0.5$. Use $L = 3.07$. Compare your results to those in part (a).

Exercises for Section 15.10

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

15.10.1 Suppose that the cost of a major repair without the extended warranty in Example 15.8 is changed to \$1000. Determine the decision selected based on the minimax, most probable, and expected cost criteria.

15.10.2 Reconsider the extended warranty decision in Example 15.8. Suppose that the probabilities of the major, minor, and no repair states are changed to 0.2, 0.4, and 0.4, respectively. Determine the decision selected based on the minimax, most probable, and expected cost criteria.

15.10.3 SS Analyze Example 15.9 based on the most probable criterion and determine the actions that are selected at each decision node. Do any actions differ from those selected in the example?

15.10.4 Analyze Example 15.9 based on the expected profit criterion and determine the actions that are selected at each decision node. Do any actions differ from those selected in the example?

Supplemental Exercises for Chapter 15

WP Problem available in WileyPLUS at instructor's discretion.
WP GO Tutorial Tutoring problem available in WileyPLUS at instructor's discretion
SS Student solution available in WileyPLUS.
VS Student video solution available in WileyPLUS.

15.S7 The diameter of fuse pins used in an aircraft engine application is an important quality characteristic. Twenty-five samples of three pins each are shown as follows:

Sample Number	Diameter		
1	64.030	64.002	64.019
2	63.995	63.992	64.001
3	63.988	64.024	64.021
4	64.002	63.996	63.993
5	63.992	64.007	64.015
6	64.009	63.994	63.997
7	63.995	64.006	63.994
8	63.985	64.003	63.993
9	64.008	63.995	64.009
10	63.998	74.000	63.990
11	63.994	63.998	63.994
12	64.004	64.000	64.007
13	63.983	64.002	63.998
14	64.006	63.967	63.994
15	64.012	64.014	63.998
16	64.000	63.984	64.005
17	63.994	64.012	63.986
18	64.006	64.010	64.018
19	63.984	64.002	64.003
20	64.000	64.010	64.013
21	63.988	64.001	64.009
22	64.004	63.999	63.990
23	64.010	63.989	63.990
24	64.015	64.008	63.993
25	63.982	63.984	63.995

a. Set up $\overline{X}$ and R charts for this process. If necessary, revise limits so that no observations are out of control.

b. Estimate the process mean and standard deviation.

c. Suppose that the process specifications are at 64 ± 0.02. Calculate an estimate of PCR. Does the process meet a minimum capability level of $PCR \geq 1.33$?

d. Calculate an estimate of PCR_k. Use this ratio to draw conclusions about process capability.

e. To make this process a 6-sigma process, the variance σ^2 would have to be decreased such that $PCR_k = 2.0$. What should this new variance value be?

f. Suppose that the mean shifts to 64.01. What is the probability that this shift is detected on the next sample? What is the ARL after the shift?

15.S8 Rework Exercise 15.S7 with $\overline{X}$ and S charts.

15.S9 WP Plastic bottles for liquid laundry detergent are formed by blow molding. Twenty samples of $n = 100$ bottles are inspected in time order of production, and the fraction defective in each sample is reported. The data are as follows:

Sample	Fraction Defective	Sample	Fraction Defective
1	0.12	11	0.10
2	0.15	12	0.07
3	0.18	13	0.12
4	0.10	14	0.08
5	0.12	15	0.09
6	0.11	16	0.15
7	0.05	17	0.10
8	0.09	18	0.06
9	0.13	19	0.12
10	0.13	20	0.13

a. Set up a P chart for this process. Is the process in statistical control?

b. Suppose that instead of $n = 100$, $n = 200$. Use the data given to set up a P chart for this process. Revise the control limits if necessary.

c. Compare your control limits for the P charts in parts (a) and (b). Explain why they differ. Also, explain why your assessment about statistical control differs for the two sizes of n.

15.S10 The following data from the U.S. Department of Energy Web site (www.eia.doe.gov) reported the total U.S. *renewable* energy consumption by year (quadrillion BTU).

Year	Total Renewable Energy Consumption (Quadrillion BTU)	Year	Total Renewable Energy Consumption (Quadrillion BTU)
1973	4.41	1995	6.56
1974	4.74	1996	7.01
1975	4.69	1997	7.01
1976	4.73	1998	6.49
1977	4.21	1999	6.51
1978	5.01	2000	6.10
1979	5.12	2001	5.16
1980	5.43	2002	5.73
1981	5.41	2003	5.94
1982	5.98	2004	6.07
1983	6.50	2005	6.23
1984	6.44	2006	6.64
1985	6.08	2007	6.52
1986	6.11	2008	7.17
1987	5.62	2009	7.60
1988	5.46	2010	8.03
1989	6.23	2011	9.00
1990	6.04	2012	8.71
1991	6.07	2013	9.27
1992	5.82	2014	9.56
1993	6.08	2015	9.58
1994	5.99		

a. Using all the data, calculate control limits for a control chart for individual measurements, construct the chart, and plot the data.

b. Do the data appear to be generated from an in-control process? Comment on any patterns on the chart.

15.S11 WP SS An article in *Quality Engineering* ["Is the Process Capable?: Tables and Graphs in Assessing C_{pm}" (1992, Vol. 4(4), pp. 563–576)] considered manufacturing data. Specifications for the outer diameter of hubs were 60.3265 ± 0.001 mm. A sample of 20 hubs was taken, and the data are shown in the following table:

Hub	x	Hub	x	Hub	x
1	60.3262	8	60.3267	15	60.3265
2	60.3262	9	60.3263	16	60.3266
3	60.3262	10	60.3269	17	60.3265
4	60.3266	11	60.3262	18	60.3268
5	60.3263	12	60.3262	19	60.3262
6	60.3260	13	60.3269	20	60.3266
7	60.3262	14	60.3261		

a. Construct a control chart for individual measurements. Revise the control limits if necessary.

b. Compare your chart in part (a) to one that uses only the last (least significant) digit of each diameter as the measurement. Explain your conclusion.

c. Estimate μ and σ from the moving range of the revised chart and use this value to estimate PCR and PCR_k and interpret these ratios.

15.S12 WP Suppose that an $\overline{X}$ control chart with 2-sigma limits is used to control a process. Find the probability that a false out-of-control signal is produced on the next sample. Compare this with the corresponding probability for the chart with 3-sigma limits and discuss. Comment on when you would prefer to use 2-sigma limits instead of 3-sigma limits.

15.S13 The following data set was considered in *Quality Engineering* ["Analytic Examination of Variance Components" (1994–1995, Vol. 7(2), pp. 315–336)]. A quality characteristic for cement mortar briquettes was monitored. Samples of size $n = 6$ were taken from the process, and 25 samples from the process are shown in the following table.

Batch	$\bar{x}$	s	Batch	$\bar{x}$	s
1	572.00	73.25	14	485.67	103.33
2	583.83	79.30	15	746.33	107.88
3	720.50	86.44	16	436.33	98.69
4	368.67	98.62	17	556.83	99.25
5	374.00	92.36	18	390.33	117.35
6	580.33	93.50	19	562.33	75.69
7	388.33	110.23	20	675.00	90.10
8	559.33	74.79	21	416.50	89.27
9	562.00	76.53	22	568.33	61.36
10	729.00	49.80	23	762.67	105.94
11	469.00	40.52	24	786.17	65.05
12	566.67	113.82	25	530.67	99.42
13	578.33	58.03			

a. Using all the data, calculate trial control limits for $\overline{X}$ and S charts. Is the process in control?

b. Suppose that the specifications are at 580 ± 250. What statements can you make about process capability? Compute estimates of the appropriate process capability ratios.

c. To make this process a "6-sigma process," the variance σ^2 would have to be decreased such that $PCR_k = 2.0$. What should this new variance value be?

d. Suppose the mean shifts to 600. What is the probability that this shift is detected on the next sample? What is the ARL after the shift?

15.S14 Suppose that a process is in control and an $\overline{X}$ chart is used with a sample size of four to monitor the process. Suddenly there is a mean shift of 1.5σ.

a. If 3-sigma control limits are used on the $\overline{X}$ chart, what is the probability that this shift remains undetected for three consecutive samples?

b. If 2-sigma control limits are in use on the $\overline{X}$ chart, what is the probability that this shift remains undetected for three consecutive samples?

c. Compare your answers to parts (a) and (b) and explain why they differ. Also, which limits would you recommend using and why?

15.S15 Consider the diameter data in Exercise 15.S7.

a. Construct an EWMA control chart with $\lambda = 0.2$ and $L = 3$. Comment on process control.

b. Construct an EWMA control chart with $\lambda = 0.5$ and $L = 3$ and compare your conclusion to part (a).

15.S16 Consider a control chart for individuals applied to a continuous 24-hour chemical process with observations taken every hour.

a. If the chart has 3-sigma limits, how many false alarms would occur each 30-day month, on the average, with this chart?

b. Suppose that the chart has 2-sigma limits. Does this reduce the ARL for detecting a shift in the mean of magnitude σ?

c. Find the in-control ARL if 2-sigma limits are used on the chart. How many false alarms would occur each month with this chart? Is this in-control ARL performance satisfactory? Explain your answer.

15.S17 WP Consider the hub data in Exercise 15.S11.

a. Construct an EWMA control chart with $\lambda = 0.2$ and $L = 3$. Comment on process control.

b. Construct an EWMA control chart with $\lambda = 0.5$ and $L = 3$ and compare your conclusion to part (a).

15.S18 WP Consider a control chart for individuals with 3-sigma limits. What is the probability that there is not a signal in 3 samples? In 6 samples? In 10 samples?

15.S19 The following data were considered in *Quality Progress* ["Digidot Plots for Process Surveillance" (1990, Vol. 23(5), pp. 66–68)]. Measurements of center thickness (in mils) from 25 contact lenses sampled from the production process at regular intervals are shown in the following table.

Sample	x	Sample	x
1	0.3978	14	0.3999
2	0.4019	15	0.4062
3	0.4031	16	0.4048
4	0.4044	17	0.4071
5	0.3984	18	0.4015
6	0.3972	19	0.3991
7	0.3981	20	0.4021
8	0.3947	21	0.4009
9	0.4012	22	0.3988
10	0.4043	23	0.3994
11	0.4051	24	0.4016
12	0.4016	25	0.4010
13	0.3994		

Construct an EWMA control chart with $\mu_0 = 0.4$, $\lambda = 0.5$, and $L = 3$ and compare to an individuals chart with 3-sigma limits.

15.S20 A process is controlled by a P chart using samples of size 100. The center line on the chart is 0.05.

a. What is the probability that the control chart detects a shift to 0.08 on the first sample following the shift?

b. What is the probability that the control chart does not detect a shift to 0.08 on the first sample following the shift, but does detect it on the second sample?

c. Suppose that instead of a shift in the mean to 0.08, the mean shifts to 0.10. Repeat parts (a) and (b).

d. Compare your answers for a shift to 0.08 and for a shift to 0.10. Explain why they differ. Also, explain why a shift to 0.10 is easier to detect.

15.S21 WP Consider the control chart for individuals with 3-sigma limits.

a. Suppose that a shift in the process mean of magnitude σ occurs. Verify that the ARL for detecting the shift is ARL = 43.9.

b. Find the ARL for detecting a shift of magnitude 2σ in the process mean.

c. Find the ARL for detecting a shift of magnitude 3σ in the process mean.

d. Compare your responses to parts (a), (b), and (c) and explain why the ARL for detection is decreasing as the magnitude of the shift increases.

15.S22 Consider an $\bar{X}$ control chart with $UCL = 32.802$, $LCL = 24.642$, and $n = 5$. Suppose that the mean shifts to 30.

a. What is the probability that this shift is detected on the next sample?

b. What is the ARL to detect the shift?

15.S23 WP SS The depth of a keyway is an important part quality characteristic. Samples of size $n = 5$ are taken every 4 hours from the process, and 20 samples are summarized in the following table.

Sample	$\bar{x}$	r	Sample	$\bar{x}$	r
1	139.7	1.1	11	138.4	0.8
2	139.8	1.4	12	138.5	0.9
3	140.0	1.3	13	137.9	1.2
4	140.1	1.6	14	138.5	1.1
5	139.8	0.9	15	140.8	1.0
6	139.9	1.0	16	140.5	1.3
7	139.7	1.4	17	139.4	1.4
8	140.2	1.2	18	139.9	1.0
9	139.3	1.1	19	137.5	1.5
10	140.7	1.0	20	139.2	1.3

a. Using all the data, find trial control limits for $\bar{X}$ and R charts. Is the process in control?

b. Use the trial control limits from part (a) to identify out-of-control points. If necessary, revise your control limits. Then estimate the process standard deviation.

c. Suppose that the specifications are at 140 ± 2. Using the results from part (b), what statements can you make about process capability? Compute estimates of the appropriate process capability ratios.

d. To make this a 6-sigma process, the variance σ^2 would have to be decreased such that $PCR_k = 2.0$. What should this new variance value be?

e. Suppose that the mean shifts to 139.7. What is the probability that this shift is detected on the next sample? What is the ARL after the shift?

15.S24 The following are the number of defects observed on 15 samples of transmission units in an automotive manufacturing company. Each lot contains five transmission units.

Sample	No. of Defects	Sample	No. of Defects
1	8	11	6
2	10	12	10
3	24	13	11
4	6	14	17
5	5	15	9
6	21		
7	10		
8	7		
9	9		
10	15		

a. Using all the data, compute trial control limits for a U control chart, construct the chart, and plot the data.

b. Determine whether the process is in statistical control. If not, assume assignable causes can be found and out-of-control points eliminated. Revise the control limits.

15.S25 WP Suppose that the average number of defects in a unit is known to be 8. If the mean number of defects in a unit shifts to 16, what is the probability that it is detected by a U chart on the first sample following the shift

a. if the sample size is $n = 4$?

b. if the sample size is $n = 10$?

Use a normal approximation for U.

15.S26 The number of visits (in millions) on a Web site is recorded every day. The following table shows the samples for 25 consecutive days.

Sample	Number of Visits	Sample	Number of Visits
1	10.12	16	9.66
2	9.92	17	10.42
3	9.76	18	11.30
4	9.35	19	12.53
5	9.60	20	10.76
6	8.60	21	11.92
7	10.46	22	13.24
8	10.58	23	10.64
9	9.95	24	11.31
10	9.50	25	11.26
11	11.26	26	11.79
12	10.02	27	10.53
13	10.95	28	11.82
14	8.99	29	11.47
15	9.50	30	11.76

a. Estimate the process standard estimation.

b. Set up a EWMA control chart for this process, assuming the target is 10 with $\lambda = 0.4$. Does the process appear to be in control?

15.S27 The following table shows the number of e-mails a student received each hour from 8:00 A.M. to 6:00 P.M. The samples are collected for 5 days from Monday to Friday.

Hour	M	T	W	Th	F
1	2	2	2	3	1
2	2	4	0	1	2
3	2	2	2	1	2
4	4	4	3	3	2
5	1	1	2	2	1
6	1	3	2	2	1
7	3	2	1	1	0
8	2	3	2	3	1
9	1	3	3	2	0
10	2	3	2	3	0

a. Use the rational subgrouping principle to comment on why an $\overline{X}$ chart that plots one point each hour with a subgroup of size 5 is not appropriate.

b. Construct an appropriate attribute control chart. Use all the data to find trial control limits, construct the chart, and plot the data.

c. Use the trial control limits from part (b) to identify out-of-control points. If necessary, revise your control limits, assuming that any samples that plot outside the control limits can be eliminated.

15.S28 An article in the *Journal of Quality in Clinical Practice* ["The Application of Statistical Process Control Charts to the Detection and Monitoring of Hospital-Acquired Infections" (2001, Vol. 21, pp. 112–117)] reported the use of SPC methods to monitor hospital-acquired infections. The authors applied Shewhart, EWMA, and other charts to the monitor ESBL Klebsiella pneumonia infections. The monthly number of infections from June 1994 to April 1998 are shown in the following table.

	Jan	Feb	Mar	April	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1994						5	0	0	2	2	3	1
1995	1	3	2	6	4	1	2	4	3	2	8	7
1996	10	6	10	11	5	6	3	0	3	3	1	3
1997	0	2	0	4	1	1	4	2	6	7	1	5
1998	3	0	1	0	2							

a. What distribution might be expected for these data? What type of control chart might be appropriate?

b. Construct the chart you selected in part (a).

c. Construct a EWMA chart for these data with $\lambda = 0.2$. The article included a similarly constructed chart. What is assumed for the distribution of the data in this chart? Can your EWMA chart perform adequately?

15.S29 An article in *Microelectronics Reliability* ["Advanced Electronic Prognostics through System Telemetry and Pattern

Recognition Methods" (2007, Vol. 47(12), pp. 1865–1873)] presented an example of electronic prognostics (a technique to detect faults in order to decrease the system downtime and the number of unplanned repairs in high-reliability and high-availability systems). Voltage signals from enterprise servers were monitored over time. The measurements are provided in the following table.

Observation	Voltage Signal	Observation	Voltage Signal
1	1.498	26	1.510
2	1.494	27	1.521
3	1.500	28	1.507
4	1.495	29	1.493
5	1.502	30	1.499
6	1.509	31	1.509
7	1.480	32	1.491
8	1.490	33	1.478
9	1.486	34	1.495
10	1.510	35	1.482
11	1.495	36	1.488
12	1.481	37	1.480
13	1.529	38	1.519
14	1.479	39	1.486
15	1.483	40	1.517
16	1.505	41	1.517
17	1.536	42	1.490
18	1.493	43	1.495
19	1.496	44	1.545
20	1.587	45	1.501
21	1.610	46	1.503
22	1.592	47	1.486
23	1.585	48	1.473
24	1.587	49	1.502
25	1.482	50	1.497

a. Using all the data, compute trial control limits for individual observations and moving-range charts. Construct the chart and plot the data. Determine whether the process is in statistical control. If not, assume that assignable causes can be found to eliminate these samples and revise the control limits.

b. Estimate the process mean and standard deviation for the in-control process.

c. The report in the article assumed that the signal is normally distributed with a mean of 1.5 V and a standard deviation of 0.02 V. Do your results in part (b) support this assumption?

15.S30 Consider the turnaround time (TAT) for complete blood counts in Exercise 15.3.12. Suppose that the specifications for TAT are set at 20 and 80 minutes. Use the control chart summary statistics for the following.

a. Estimate the process standard deviation.

b. Calculate PCR and PCR_k for the process.

15.S31 An article in *Electric Power Systems Research* ["On the Self-Scheduling of a Power Producer in Uncertain Trading Environments" (2008, Vol. 78(3), pp. 311–317)] considered a self-scheduling approach for a power producer. The following table shows the forecasted prices of energy for a 24-hour time period according to a base case scenario.

Hour	Price	Hour	Price	Hour	Price
1	38.77	9	48.75	17	52.07
2	37.52	10	51.18	18	51.34
3	37.07	11	51.79	19	52.55
4	35.82	12	55.22	20	53.11
5	35.04	13	53.48	21	50.88
6	35.57	14	51.34	22	52.78
7	36.23	15	45.8	23	42.16
8	38.93	16	48.14	24	42.16

a. Construct individuals and moving-range charts. Determine whether the energy prices fluctuate in statistical control.

b. Is the assumption of independent observations reasonable for these data?

15.S32 Consider the TAT data in Exercise 15.3.12.

a. Construct an EWMA control chart with the target equal to the estimated process mean, $k = 0.5$, and $h = 4$. Does the process appear to be in control at the target?

b. If the mean increases by 5 minutes, approximate the chart's ARL.

15.S33 Consider the infectious-waste data in Exercise 15.4.6. Use the data after the process change only.

a. Construct an EWMA control chart with the target equal to the estimated process mean, $k = 0.5$, and $h = 4$. Does the process appear to be in control at the target?

b. If the mean increases by 1.0 lb, approximate the ARL of the chart.

Index

Note: A page number followed by an *f* refers to a figure (graph, illustration); a *t* refers to a table.

1/4 fraction, called a 2^{k-2} fractional factorial design, 418
1/8 fraction, called a 2^{k-3} fractional factorial design, 418
1/16 fraction, called a 2^{k-4} fractional factorial design, 418
2^2 factorial design, 390–391, 390*f*, 391*t*
  interaction effect, 391
  main effects of factors, 391
  models and residual analysis, 401
  signs for effects in the, 391*t*
2^3 factorial design, 396, 396*f*, 397*f*
  main effects of factors, 396
  three-factor interaction effects, 398, 398*t*
  two-factor interaction effects, 397
2^k factorial design, 390–402
  addition of center points to a, 405–407, 406*f*
  blocking and confounding in the, 408–413, 408*f*, 409*f*
  models and residual analysis, 393
  one-half fraction of the, 413–418, 413*t*, 414*f*
  projection of, 401
  response and residuals by fitting a regression model, 393–394
2^{k-1} factorial design, projection of the, 417
2^{k-2} fractional factorial design, called a 1/4 fraction, 418
2^{k-3} fractional factorial design, called a 1/8 fraction, 418
2^{k-4} fractional factorial design, called a 1/16 fraction, 418
2^{k-p} fractional factorial, 418–424
2^{k-p} fractional factorial design, 422, 423*t*–424*t*
3-sigma control limits, 460
6-sigma process, 455, 455*f*

A

Abscissa scale, 249
Abstraction, also called a mathematical model, 18
Acceptance region, 196
Addition rule, 30
Adjusted R^2 statistic, 324
Adjustments, when they should and should not be made to a process, 10–11, 11*f*
Agresti-Coull confidence interval, 188
Alias pattern, fractional factorial design and, 417
Alias structure, 419*t*
  how to find the, 422
Aliases, definition of, 414
All-possible regressions approach, 349
  fit all regression equations for one candidate variable, 342
Alternate fraction, 414, 414*f*
Alternative hypothesis, 195, 206, 255, 258, 262, 265, 271, 275
  abscissa scale of the operating characteristic curve and, 249
  one- or two-sided, 249, 250, 259, 276
  part of the procedure for hypothesis tests, 206
Analysis, regression, 2
Analysis of variance (ANOVA), 354–359, 354*t*, 382–383. *See also* ANOVA
  for a single-factor experiment, fixed-effects model, 357*t*
  for first-order model, 430*t*
  for quadratic model, 430*t*
  for regression model, 320
  to test for significance of regression, 291–292
Analysis of variance identity, equation for, 291
Analysis of variance method, 366
Analysis, of variance table, 291, 291*t*
Analytic study, definition of, 11
  versus enumerative study, 11, 12*f*
ANOVA
  computing formulas, randomized block experiment, 371
  F-test, 357
  sum of squares identity, single factor experiment, 356
  sum of squares identity, two factors, 383
  sum of squares identity, randomized complete block experiment, 370
  variance components and, 365–368
  variance components estimates, 367
ANOVA table, for a two-factor factorial, fixed-effects model, 385*t*
ANOVA test, hypothesis for, 322
ANOVA test statistic, 322
Approximate confidence interval
  on a binomial population, 186
  on difference in population proportions, 277
Approximate sample size
  for a one-sided test on a binomial proportion, 228
  for a two-sided test on a binomial proportion, 228
  for a two-sided test on the difference in population proportions, 276
Approximate sampling distribution, of a difference in sample means, 156
Approximate tests
  on a binomial proportion, 226
  on the difference of two population proportions, 274
Approximate type II error
  for a one-sided test on difference of two population proportions, 276
  for a two-sided test on difference of two population proportions, 276
Approximation
  large-sample, 263
  Wilcoxon rank-sum test and normal, 263
Arbitrary normal random variable, how to calculate probabilities for, 77
Architecture, advantage and disadvantage of parallel, 38
Area, Poisson distribution and, 64
Arithmetic mean, known as the sample mean, 127
ASMT Standard E23, 171
Assignable causes, definition of, 437
Assumptions
  for two-sample inference, 246–247
  needed to fit a regression model, 296
  verifying, 144
Asymptotic normal distribution, 166
Attribute control charts, 439, 456–460
Average, mean and variance of an, 119
Average run length (ARL), 460–461, 461*t*
Axioms
  allow easy calculation of probability, 29
  do not determine probability, 28
Axioms of probability, 26, 28
  definition and equation for, 29

B

β-error, type II error, 198
Backward elimination algorithm, for candidate regressors, 347–348, 348*t*
Bayes, Thomas, 39

Bayes estimator, 168–169
Bayes' theorem
  equation for, 39–40
  to calculate conditional probabilities, 39
Bayesian approach, 167
  parameter estimation and the, 161
Bayesian estimation, of parameters, 167–169
Bernoulli distribution, maximum likelihood estimator and, 163
Bernoulli random variable, 305
Bernoulli trial
  binomial distribution and, 52
  binomial random variable and, 59
  definition of, 51, 5
  geometric random variable and, 55, 56*f*
  negative binomial distribution and, 57–59
  subintervals and, 63
Beta distribution, 92–93
  mean and variance of a, 93
Beta probability density functions, for selected values, 93*f*
Beta random variable, 92, 93
Bias, of an estimator, 156
Biased estimator, 157, 160, 161*f*
Bimodal distribution, negative, 110
Bin, range of data in a frequency distribution, 135
Binomial coefficient, 53, 53*f*
Binomial distribution, 51–55, 58, 63, 79
  as an approximation to the hypergeometric distribution, 81
  Bernoulli trial and, 52
  common joint distributions and the, 113–114
  compared to hypergeometric distribution, 62, 62*f*, 62*t*
  definition of, 52–55, 53*f*
  negative, 118
  normal approximation and, 79, 80, 80*f*
  when it is not symmetrical, 82*f*
Binomial expansion, equation for, 53
Binomial experiment, 54, 114
Binomial model, 79
Binomial population
  approximate confidence interval on a, 186
  sample size for a specified error on a, 187
Binomial probabilities, can be approximated with probabilities based on normal distribution, 81
Binomial probability, 79
Binomial probability mass function, 54
Binomial proportion
  approximate sample size for a one-sided test on a, 228
  approximate sample size for a two-sided test on a, 228
  approximate tests and, 226
  normal approximation for a, 185
  small sample tests on a, 227
  testing hypothesis and, 226
Binomial random variable, 54, 80
  as opposite of negative binomial random variable, 59
  Bernoulli trial and, 59
  definition of, 52
  mean and variance of a, 54–55
  moment-generating function for a, 123
Bins
  how to choose the number of, 135
  unequal width of, 136
Bits, digital communication channel and transmitted, 31
Bivariate distribution, 97
Bivariate normal distribution, 115, 115*f*
  probability density function of a, 115
Bivariate normal probability density function, 116*f*
Bivariate normal random variable
  correlation of, 116
  marginal or conditional distribution and, 116
  zero correlation implies independence, 117
Bivariate probability distribution, 97
Bivariate probability function, 115
Block, principle, 409
Blocking
  in the 2^k factorial design, 408–413, 408*f*, 409*f*
  when is it necessary, 372
Blocks, confounded with, 408
Bonferroni confidence intervals, for standard deviations, 272
Bootstrap confidence interval, 189
Bootstrap samples, 160
Bootstrap standard error, computer-intensive technique, 159–160
Bound, one-sided upper- or lower-confidence, 252
Bounds
  lower- and upper-confidence limits, 173, 176
  one-sided confidence, 176, 181, 187
Box plot
  also called a box-and-whisker plot, 139, 139*f*
  type of graphical display, 139, 139*f*
Box-and-whisker plot, 182*f*
Business cycle, variability and a, 140

C

C, catalyst, 12
C chart, produced by computer, 460
Candidate regressors, 347
Candidate variables, 342, 348
Capability, actual and potential, 454
Categorical data, 138
Categorical regressors, indicator variables and, 339–341
Categorical variable, 339
Causal sense, 284
Causality, in observation studies is difficult to identify, 246
Cause-and-effect diagram, 472, 472*f*, 476
  an SPC problem-solving tool, 436
Cause-and-effect relationship, 6, 246
Cell, range of data in a frequency distribution, 135
Census, definition of, 5
Center line (CL)
  location of upper and lower control limits from the, 11
  meaning of a control chart's, 11
  of a control chart, 437, 437*f*
Center points, 405, 417
Central composite design (CCD), 431–432, 432*f*
  axial spacing and rotatable, 432, 432*f*
Central limit theorem, 152–154, 153*f*, 153*t*
    computer simulation experiment demonstrating, 152, 153*f*, 153*t*
  definition of, 15
  of De Moivre, 73–74
  sample size large enough so can apply, 154
  sampling distributions and, 150–151
Central tendency, of the data, 127, 133, 136
Chance causes, 437
Charpy V-notch (CVN) method, for notched bar impact testing of metallic metals, 171
Chart. *See also* Control chart
  NP, 458
  Pareto, 138, 476
  tier, 452
  time-weighted, 462–471
  tolerance, 452
Chart, *U*, 458
Chebyshev's inequality, 159
Check sheet, an SPC problem-solving tool, 436
Chi-square
  random variable, 183
  upper and lower points of, 183
Chi-square (χ^2) distribution, 89, 145, 183–184
Chi-square test
  operating characteristic curves for, 224
  P-value of a, 223
Circuit
  advanced, 38, 38*f*
  parallel, 38, 38*f*
  series, 38, 38*f*
Circumflex, 13
Class interval, range of data in a frequency distribution, 135

Coefficient
binomial, 53, 53*f*
confidence, 173
correlation, 299
of multiple determination R^2, 324, 349
partial regression, 312
Pearson correlation, 143
regression, 284, 321, 325
relationship between an effect and a, 394
regression, 281, 321
sample correlation, 143, 300
standard error of a, 394
Coefficient of determination (R^2), 298–299
used to judge adequacy of regression model, 298
Coefficient of restoration, 219, 219*f*
Column average, 386
Combinations
and subsets, 25
definition and equation for, 25–26
Common joint distributions, examples, of, 113
Communications channel, corrupted pulse in a digital signal over a, 26, 27*f*
Communications network, phone lines and information capacity, 19
Comparative experiment, 7
objective of, 246
Complement
definition and notation for, 21
of an event, 21, 23, 29
type of set operation, 21
Complete defining relation, 419
Completely randomized design (CRD), 355, 382
Completely randomized experiment, 246
Computer
C chart produced by, 460
calculating one-sample T, 220
calculating power and sample size for a 1-sample *Z*-test, 214–215
calculating sample sizes and type II error probabilities, 214–215
calculating test and CI for one proportion, 227
calculating test and CI for one variance, 224
calculations for test for equal variances, 272
calculations for two-sample hypothesis test and CI procedure for proportions, 275
calculations for two-sample *t*-test and confidence interval, 256*f*, 258
construction of an *R* control chart and $\overline{X}$ control chart, 449
important tool in data analysis, 127
usually used in fitting multiple regression models, 319
Computer software, available to compute distributions, 145
Computing formulas for ANOVA, single factor with unequal sample sizes, 359
Concentration, input feed stream, 6
Concept, of equally likely outcomes, 27
Conceptual model, of repeated replications, 26
Conceptual population, 128
Conditional distribution, of bivariate normal random variables, 116
Conditional mean and variance, 104
Conditional probability, 31–34, 32*f*, 39
independence and, 36
multiplication rule and, 34
of B given A, 31
of event B given an event A, 33
of parts with surface flaws, 32*f*
random samples and, 34
Conditional probability density function, 102–104, 103*f*
Conditional probability distribution, 102–106, 109
Conditional probability mass function
for continuous random variables, 102–103, 103*f*, 105
Conditional variance, 104
Confidence bounds, binomial population and approximate one-sided, 187
Confidence coefficient, 173
Confidence interval (CI), 292–294
Agresti-Coull, 188
Bonferroni, 272
bootstrap, 189
definition of, 171
development and basic properties of, 172–173
for a correlation coefficient, 301
for μ_D from paired samples, 267
guidelines for constructing, 188–189, 189*t*
how to construct, 171
how to interpret, 174, 174*f*
hypothesis testing and relationship to, 206
in multiple linear regression, 329–331
large-sample for mean, 177
large-sample, 251
measure of reliability of a procedure, 171
method to derive a, 176–177
on a difference in treatment means, 359
on a treatment mean, 358
on a variance, 184
on individual regression coefficients, 329–330
on mean response, 330–331
on parameters, 293
on slope and intercept, 292–293
on the difference in means, variances known, 251–252
on the difference in means, variances unknown and equal, 260
on the difference in means, variances unknown and not assumed equal, approximate, 261
on the difference in population proportions, 277–278
on the mean of a normal distribution, variance known, 172–179
on the mean of a normal distribution, variance unknown, 179–182
on the mean response, 293–294
on the mean, variance known, 173
on the mean, variance unknown, 181
on the ratio of variables from two normal distributions, 273
parameter and large-scale, 179
purpose of, 172
t, 181
two-sided, 181, 186
used in engineering and science, 172
Confidence level, precision of estimation and, 174–175, 175*f*
Conformance, quality of, 435
Confounded, with blocks, 408
Confounding, in the 2^k factorial design, 408–413, 408*f*, 409*f*
Contingency table tests, 232–234, 232*t*
Continuity correction, 79, 80, 82
Continuous cumulative distribution function, 70
Continuous probability distribution, known as normal distribution, 67
Continuous random variable
conditional probability mass function and, 102–103, 103*f*
defined as having a continuous cumulative distribution function, 70
definition and examples of, 41, 67
distribution of as described by cumulative distribution function, 70
expected value of a function of a, 72
general functions of a, 121
mean and variance of a, 71
probability density factor and a, 70
probability density function and, 71
probability distributions and, 66–94
standard deviation and, 71
Continuous sample space, definition of, 20
Continuous uniform distribution, 72–73
probability density function and, 72, 72*f*
Continuous uniform random variable, 72
mean and variance of, 73
Contour plot, 115, 115*f*
Contrast, relationship between an effect and a, 392
Contrasts, 391

Control chart. *See also* Chart
analysis of patterns on, 442–444, 442*f*
attribute, 439, 456–460
center line (CL), 437, 437*f*
cumulative sum, 452
CUSUM, 468
design of a, 440–441
EWMA, 463
exponentially weighted moving-average (EWMA), 452, 462–465
for averages ($\overline{X}$ control chart), 444–446
for chemical concentrations, 12*f*
for fractional nonconforming (*P* chart), 456
for individual measurements, 450–452
fraction-defective, 456
general model for a, 438
individuals (*X* chart), 450
introduction to, 436–443
meaning of the center line on a, 11
most powerful SPC problem-solving tool, 436
pattern recognition on a, 443
prevents unnecessary process adjustments, 440
primary tool to understand process variability, 435
process improvement and detect assignable causes, 438, 439
provides diagnostic information, 440
provides information about process capability, 440
range (*R*), 444–445
Shewhart, 438, 460
standard deviation (*S*), 444–446
time-weighted, 452
to look at variability in time-oriented data, 11
total defects in a sample (*C* chart), 460
two types of, 439
upper control limit (UCL) and lower control limit (LCL) on a, 437, 443
used as an estimating device, 439
used for controlling/improving a process, 12, 12*f*
used for defect prevention, 439
used for online process monitoring, 438
used to improve productivity, 439
variables, 439
$\overline{X}$, 440, 440*f*, 442*f*
Control chart performance, 460–461
Control limit
can help decide actions to adjust/improve the process, 11, 15
for CUSUM, 467
lower, 11, 12*f*
upper, 11, 12*f*
Cook's distance formula, 336
Cook's distance measure, 336
Correction factor
finite population, 62
used to improve approximation, 81
Correlated, definition of, 113
Correlation, 299–302
nonzero, 113
of bivariate normal random variables, 116
relationship between two random variables, 112–113
zero, 301
Correlation coefficient, 299
confidence interval for a, 301
Correlation model, 280
Countably infinite, 20
Counting techniques, 23–26, 21*f*
Covariance, 110–111
defined for both continuous and discrete random variables, 111
joint probability distribution and the sign of, 111*f*
measure of linear relationship between random variables, 111, 111*f*
measure of relationship between two random variables, 110
positive or negative, 111*f*
sign of, 111*f*
zero, 111*f*
Covariance matrix, 321
C_p statistic, 342
Criterion
expected cost, 475
minimax and minimin, 474
most probable, 472
Critical region, 196, 198*f*, 210, 222*f*
definition of, 196, 198*f*
for one- and two-sided alternative hypotheses where to place the, 209, 209*f*
for one- or two-sided alternative hypothesis, 240
for one-sided alternative hypothesis, 226
locate it to control the type I error, 216
location (two-tailed, upper-tailed, lower-tailed) of the, 206, 209*f*
or rejection region, 210
what determines the location of the, 218
Critical values, boundaries between critical and acceptance regions, 196
Cumulative distribution, of a standard normal random variable, 75
Cumulative distribution functions (CDF), 45–47, 70, 83, 84, 208
defined at noninteger values, 46
equation for, 45–46, 70
for distribution of discrete random and continuous random variables, 70
of a discrete random variable, 45
often used to compute probabilities, 89
probability mass function and, 46
properties of, 46
Cumulative distribution plot, 137*f*
Cumulative frequency plot, variation of histogram, 137
Cumulative Poisson probabilities, 88
Cumulative probabilities
and probability mass function, 45
used to describe random variable's probability distribution, 45
used to find probability mass function of a discrete random variable, 45
Cumulative sum control chart (CUSUM), 465–467, 466*t*, 467*f*. *See also* CUSUM
as a time-weighted chart, 452
Current source drift, 18
Curvature sum of squares, *t*-statistic and, 406
Curvilinear regression, 357
Curve
bell-shaped probability density function, 74
probability and area under a normal, 80*f*
using operating characteristic (OC), 213–214
CUSUM, how to devise control limits for, 467
CUSUM control chart, 468

D

Data
categorical, 138
central tendency of the, 127, 133, 136
collected to evaluate current or future production, 11
collecting engineering, 5–12
copies written to multiple drives, 43
Deming and different conclusions from, 11
difference between a census and sample, 5
distributed among multiple disks, 43
dot diagram used for limited amount of, 4
historical archived, 5, 6
in the form of ranks, 234
location of on a dot diagram, 4
methods of collecting, 5–7
multivariate, 138, 142
numerical description of features of, 127
numerical summaries of, 127–131
parity, 43
plotting the, 127
retrospective study and use of historical, 5
scatter of on a dot diagram, 4
statistics is the science of, 3
time-to-failure, 144

transcription or recording error that affects, 6
unimodal, 137
use control chart to look at variability in time-oriented, 11
useful information versus, 6
variability of on a dot diagram, 4
wire bond pull strength, 14*t*
De Moivre, central limit theorem of, 73–74
Decision criteria, 474–476
for testing a hypothesis, 196*f*, 197*t*
Decision interval, 468–470
Decision making, statistical inference is used in, 149
Decision models, definition of, 473
Decision theory, 473–476
Decision tree, 475, 475*f*
Decisions, statistical inference and, 150
Defect concentration diagram, 472, 472*f*, 476
an SPC problem-solving tool, 436
Defective parts
and nondefective parts, 26
surface flaws and, 32, 32*t*
Defects, quality improvement and elimination of, 435
Defects-per-unit control chart (*U* chart), 458
Defining contrast, 409
Defining relation, for the design, 413
Degrees of freedom, 218
definition of, 130
Deming, W. Edwards, 9
14 points for quality and productivity improvement, 477–478
analytic study and, 11
bowl of beads experiment, 435
control charts and, 11
funnel experiment of, 10*f*
industrial statistician, 9
stressed the variability of processes over time, 9
use of data for different types of conclusions, 11
DeMorgan's laws, 23
Density function, used in engineering to describe physical systems, 67
Dependent variable, 312
Descriptive statistics, 126–147
definition of, 127
Design
completely random, 382
of an experiment, 376
Design and analysis, of single-factor experiments, 351–374
Design and statistical analysis, 368–372
Design generators, 419
Design matrix, 396
Design of experiments, with several factors, 374–433
Design quality, 435
Design resolution, to catalog fractional factorial designs according to alias patterns, 417
Designed experiment, 5, 6–7, 7*f*
deliberates changes made to the controllable variables, 6
is employed sequentially, 353
only way to determine cause-and-effect relationship, 284
used to study complex systems with several factors, 7
Designing engineering experiments, 352–353
Deterministic linear relationship, 281
Deviation, standard, 71
Diagram
cause-and-effect, 472, 472*f*, 476
defect concentration, 436, 472, 472*f*, 476
describes relationship between events, 22
dot, 4, 128, 128*f*
ordered stem-and-leaf, 133
Pareto, 471, 471*f*
portrays relationship between sets, 22
scatter, 142–144, 143*f*
several mutually exclusive events and a Venn, 30, 31*f*
stem-and-leaf, 131–132, 132t, 133*f*
tolerance, 453*f*
tree, 21*f*, 23–24, 33, 33*f*
Venn, 22, 23*f*, 35
Die height, 13, 13*f*, 14*f*
Digidot plot, invented by Hunter, 140–141, 141*f*
Discrete random variable
definition and examples of, 41, 43
definition of the simplest, 49
distribution of as described by cumulative distribution function, 70
equation for expected value of function of a, 49
general functions of a, 120
mean and variance of a, 47
probability distributions and, 42–65
with a countably infinite range, 45
Discrete sample space
definition of, 20
probability for, 26
Discrete uniform distribution, 49–51
equation for, 49
Discrete uniform random variable, mean and variance of, 50–51
Dispersion, 50
variability and, 47
Distillate, 5
Distillation column, 6–9, 7, 7*f*, 8*f*
acetone-butyl alcohol, 5, 7*f*
designed experiment for the, 7*t*
four-factorial experiment for a, 8, 8*f*
retrospective study of, 5–6
Distributed rule, for set operations, 30
Distribution
asymptotic normal, 166
beta, 92–93
bimodal, negative, 110
binomial, 51–55, 58, 79, 113–114
bivariate, 97
bivariate normal, 115
bivariate probability, 97
chi-square (χ^2), 89, 145, 182–184
conditional, 116
conditional probability, 105, 109
continuous probability, 67
continuous uniform, 72–73
cumulative, 45–47, 70, 83, 84, 208
Erlang, 86–87
exponential, 83–84
F, 268–270, 269*f*
frequency, 135–136, 135*t*
gamma, 145
Gauss developed the normal, 73
Gaussian, 73
geometric, 55–57, 56*f*
heavy-tailed, 146, 147*f*
hypergeometric, 59–63, 61*f*, 81
joint probability, 107–110, 111*f*
light-tailed, 146, 147*f*
lognormal, 91
marginal, 116
marginal probability, 99, 100*t*, 109, 109*f*
mean and sampling, 151
multinomial probability, 113–114
negative binomial, 57–59, 110, 118
noncentral *F*-, 363
nonsymmetrical, 92
normal, 67, 73–79, 75*f*, 77*f*
null, 216
of the ratio of sample variances from two normal distributions, 270
Poisson, 64–65, 64*f*, 81, 230
posterior, 167
prior, 167
probability, 44, 67, 67*f*, 167
Raleigh, 89
random variable and its individual probability, 99
reference, 209, 216, 247
relative frequency, 135
sample means and approximate sampling, 156
sampling, 149
symmetric continuous, 239
symmetrical and skewed histogram, 138*f*
t, 180, 180*f*
uniform, 92
Weibull, 86, 89–90
width of a normal, 75
χ^2, 182–183, 183*f*
Distribution-free methods, parametric and, 234
Distribution function, cumulative (CDF), 45–47, 70, 83, 84, 208
Distribution moment, 162

Dodge, Harold F., 436
Dot diagram
  data location and scatter on a, 4
  of pull-off force data, 4*f*
  physical interpretation of sample mean, 128, 128*f*
  pros and cons of a, 4
  shows variation but not cause of problem, 9, 9*f*
  used for limited amount of data, 4
Drift, current source, 18
Ductal-to-brittle transition, 171
Dummy variable, 340

E

ϵ, definition of, 12
Effect
  relationship between a coefficient and an, 394
  relationship between a contrast and an, 392
Effects, normal probability plot of, 417
Efficiency, loss of in nonparametric procedure, 234
Electrical current, example of continuous random variable, 41
Electrical science, 3
Elimination, quality improvement and waste, 435
Empirical model, 281–284
  definition of, 13
  plot of predicted values of pull strength, 14*f*
Engineer, problem-solving and the, 2
Engineering, role of statistics in, 1–16
Engineering method, steps in the, 2–5, 3*f*
Enumerative study
  definition of, 11
  versus analytic study, 11, 12*f*
Equality claim, null hypothesis and, 195
Equally likely outcomes, 27
Equivalence band, 241–242
Equivalence interval, 241
Equivalence testing, 240–242
Erlang distribution, 86–87
Erlang random variable, 87, 88
  mean and variance of an, 118
Error, sampling, 5
Error in estimation, 175
Error of estimation, 190
Error mean square, 299, 324–325, 335, 356
Error propagation, 118–119
Error sum of squares, 291
  computer software calls it "residual" sum of squares, 292
  definition of, 356
  sum of squares of residuals and the, 287
Error variance, 283
Estimated standard error, 159, 321
  definition of, 288
Estimating, variance of the error term ϵ, 287
Estimator
  asymptotically unbiased, 166
  Bayes, 168–169
  bias of an, 156
  biased, 157, 160, 161*f*
  exponential distribution moment, 162
  gamma distribution moment, 163
  least squares, 288, 321
  logical point, 247
  maximum likelihood, 163, 165–166, 179
  mean squared error of an, 160–161
  minimum variance unbiased (MVUE), 158
  moment, 162
  normal distribution moment, 162
  of variance, 287
  of variance, equation for, 320
  point, 161–169, 295
  sampling distribution of two unbiased, 158*f*
  standard error of an, 158
  standard error of the point, 185
  unbiased, 156–158, 158*f*, 287–288
  variance and a pooled, 253–254
Event, 21–23. *See also* Events
  as a subset of the sample space, 21
  calculating the probability of an individual, 29
  definition of, 11
  definition of the complement of an, 23
  exhaustive, 36
  mutually exclusive subsets and partitioning of an, 36, 36*f*
  mutually exclusive subsets by partition of an, 31
  rare, 155
  sample, 18–23
Events
  complement of, 29
  diagram describes relationship between, 22
  intersection of, 29, 34–36
  joint, 29
  mutually exclusive, 23
  mutually exclusive relationship between two, 37
  union of, 29
  used to define outcomes of interest, 22
EWMA control chart, 463
EWMA update equation, 463
Example
  access time of computer server (milliseconds), 99, 99*f*, 101, 101*f*
  acetone-butyl alcohol distillation column, 5, 7*f*
  acetone concentration readings taken hourly, 7–9, 12*f*
  adhesion properties of different aircraft primer paints, 385–386, 385*t*, 386*f*, 386*t*, 387*f*, 387*t*
  advanced circuit, 38, 38*f*
  air crew escape system and burning rate of solid propellant, 194–198, 201–203. 207, 211, 213–214
  air quality monitoring stations, 96
  air samples containing a rare molecule, 51
  airplane sidewall panels cost per unit, 338, 338*f*, 338*t*, 339*t*
  alloy adhesion of U-700 alloy specimens load at failure, 181, 182*f*
  alloy compression strength, 132
  annual sales for a company over time, 140, 140*f*
  arsenic concentration in public water, 257–258, 257*f*
  arsenic contamination of drinking water, 245
  as a discrete random variable, 43
  as a lognormal random variable, 91
  asbestos particles in a square meter of dust, 82
  astaxanthin (a carotenoid) commercial production, 376
  automated filling of bottles with liquid detergent, 223, 224
  automatic machine used to fill bottles with detergent, 184
  automobile crankshaft bearings, 186, 187
  average molecular weight of a polymer, 12
  battery life in a portable personal computer, 145
  calculations for number of flaws in thin copper wire, 64
  camera flash recycle time of cell phone tested until fails, 20, 22
  camera flashes time to recharge and third failure, 59
  cell phone camera flash recycle time, 20, 22
  cement hydration with standard calcium and cement doped with lead, 260–261
  central computer processor failure, 86
  ceramic substrate production line, 457, 457*f*, 457*t*
  chance of rain today, 26
  chemical batch yield, 134, 134*f*
  chemical process concentration and hourly output, 451, 451*f*, 451*t*
  chemical process concentration EWMA, 464, 464*f*, 465*t*
  chemical process concentration tabular CUSUM, 468–469, 469*t*, 470*f*, 470*t*
  chemical process yield steepest ascent, 427, 427*f*

coin flips and number of heads obtained, 51
comparing tips for a hardness-testing machine, 264
components with a standard mean resistance of ohms, 240–241
compressive strengths of aluminum-lithium alloy, 131, 132*t*, 133*f*, 135*t*
computer usage and user log-ons, 84
continuous sample space and low/medium/high camera recycle time, 20
corrupted pulse in a digital signal over a communications channel, 26, 27*f*
curing methods on mean compressive strength of concrete, 352
current in a thin copper wire, 12
current in a thin copper wire over time, 18, 19, 19*f*
cycle time to process loan applications, 160
density of loading on long, thin beam, 67–68, 67*f*
development of contract decision problem, 476
different formulations for a soft-drink beverage, 234–235
digital channel and bit errors, 45
digital channel and four bits received in error, 57
digital channel and number of next four bits received, 52, 52*t*
digital channel and numbers of errors in next four transmitted bits, 49
digital channel and transmission error with fifth bit, 55
digital channel of bits received, 113–114
digital communication channel error rate, 31
digital transmission channel and all bits received are errors, 51
digital transmission channel and bit received in error, 44, 44*f*
dimensional length on a manufactured part, 67
discrete sample space of yes/no camera recycle time specifications, 20
discrete sample space that is countably infinite, 20
distal communications channel bits received in error, 80
distance between flaws in a length of copper wire, 83
distribution of throws of single, six-sided true dice, 154, 154*f*
door hinge on an automobile, 377
eating different types of chocolate and cardiovascular health, 258–259
electrical current and power with resistance in ohms, 72
electrical current in a thin copper wire, 69, 70, 71
electrical current in electronic manufacturing, 454
electrical current measurements (milliamperes) in a wire, 74, 74*f*
electronic assembly and lifetime of four components, 107–108
electronics company resistors, 155, 155*f*
emergency department (ED) wait times, 2
emergency department electrocardiogram results, 18
epitaxial process on polished silicon wafers, 392–393, 392*t*, 393*t*, 395*f*
etch variability in thickness of oxide layers on semiconductor wafers, 271
expected value of a function of two random variables, 110
extended warranty decision problem, 473, 473*t*
fabric strength and effect of different chemicals, 371, 372*t*
fabric tensile strength and loom-to-loom variability, 367, 367*t*, 368*f*
factors that affect gasoline mileage, 3
flash recharge time in cell-phone camera, 43, 43*t*
gasoline mileage, 138
gasoline mileage performance of a vehicle, factors that affect, 311
Geiger counter detection of a particle, 85
golf club design and coefficient of restitution, 219, 219*f*, 221
grinding process on surface finish of impellers in jet-turbine engines, 273
hardwood concentration and Fisher LSD method, 360, 360*f*, 361*f*
health insurance plan preference, 232, 232*t*
heights of fathers and sons, 283
high iron levels and increase in heart attack, 246
hole diameter drilled in metal sheet component, 69
hospital emergency department (ED), probability model and example of, 2
hospital emergency room visits and outcomes, 22, 22*t*
hospital operating schedule, and sequencing of surgeries, 25
injection-molded part, 117
injection-molding and excessive shrinkage, 419–422, 420*t*, 421*f*, 422*f*
is modeled as a discrete random variable, 43
joint and marginal probability distribution, 105
jury trial, 201
laser diode in a batch and power requirements, 27, 27*f*
lead-free solder in surface mount technology (SMT), 376–377, 377*f*
length and width of a manufactured part, 120
length of one dimension of an injection-molded part, 96, 98, 98*f*
lengths of two dimensions of a machined part, 106, 107
lifetime of a semiconductor chip, 86
loading on long thin beam puts mass at discrete points, 44, 44*f*, 47, 47*f*, 48*f*
machining stages for high-rpm pistons, 35
manufactured defective and nondefective parts, 26
manufactured parts and nonconforming parts, 59–60
manufacturing inspection of defective and nondefective parts, 26, 28
marketing and new product design, 48–49
mean axial stress in tensile members in aircraft, 262–263, 262*t*, 263*t*
mean fill volume of a soft drink container, 150
mean purity of oxygen from hydrocarbon, 283, 283*f*
mean viscosity of a chemical product, 171–172
mean yield from catalyst sample size, 259–260
means being compared in completely randomized experiment, 365, 365*t*
mechanical shaft bearing with wear and time to failure, 90
medical diagnostic and screening of population, 40
mercury contamination of largemouth bass, 178, 178*f*
metallic metal transition and CVN test, 175
missile miss distance of shoulder-fired ground-to-air missile, 410–412, 410*f*, 411*f*, 411*t*, 412*t*
mobile response time, 102
model of an exponential random variable, 86
modeled as a Poisson process, 86
modeled by a binomial random variable, 80
most popular model of Boeing transport aircraft, 138, 138*f*
multiple-choice test and questions answered correctly, 51

Example (*contd.*)
nitride etch process on single-wafer plasma etcher with one-half fraction, 416, 416*f*, 416*t*
nitride plasma etch process on single-wafer plasma etcher, 402–405, 402*t*, 403*f*, 403*t*, 404*t*, 405*f*
normal distribution current measurement in a wire, 78, 78*f*
normally distributed current in a wire and measurement, 79, 79*f*
notebook computer power supply distribution, 231–232, 231*t*
number of female births out of number of hospital births, 51
number of signal bars and response time, 112
number of voice lines in use, 50
nylon connector for automotive engine, 4, 6, 7
of a 2^2 factorial design, 392*t*
of a binomial coefficient, 53, 53*f*
of a comparative experiment, 7
of a designed experiment, 6, 7
of a negative binomial distribution, 57
of a Poisson distribution, 82
of a Poisson process, 64, 84, 88
of a probability model, 15
of a regression model, 13
of a retrospective study, 5–6
of a tree diagram and conditional probabilities, 33, 33*f*
of a Weibull distribution, 91
of an analytic study, 11
of an empirical model, 13, 14*f*
of an enumerative study, 11
of analysis of variance, 292
of ANOVA method, 371, 372*t*
of Bayes estimator for the mean of a normal distribution, 168
of Bernoulli distribution MLE, 163
of binomial probability mass function, 54
of binomial random variables, 54
of bivariate normal distribution, 115, 117
of bootstrap standard error, 160
of box-and-whisker plot, 182
of central limit theorem, 155, 155*f*
of concept of relative frequency, 27*f*
of condition probability, 32–33
of conditional mean and variance, 104
of conditional probability, 31, 102, 103
of conditional probability distribution, 105
of continuous random variables, 92
of control chart for chemical concentration, 12*f*
of cumulative distribution function, 46, 46*f*, 70*f*, 71
of designing engineering, 352–353
of distribution of a sum of Poisson random variables, 124
of error propagation, 119
of exponential distribution MLE, 164
of exponential distribution moment estimators, 162
of Fisher's LSD method, 372, 373*t*
of fraction-defective control chart (*P* chart), 457
of gamma distribution moment estimators, 163
of geometric random variable, 56
of histogram, 137*f*
of hypergeometric distribution simplifies calculations, 61
of hypergeometric random variable, 60
of joint density function, 116, 116*f*
of joint probability density function, 99
of Karlsruhe and Lehigh procedures, 265–266
of lack of memory property, 57, 85
of linear function of independent normal random variables, 120
of linear relationship, 113
of marginal probability distribution, 109, 109*f*
of matrix of scatter plots, 343*f*
of maximum likelihood estimators of mean and variance, 166
of mean and variance, 62
of mean resistance and standard deviation, 155, 155*f*
of moment-generating function for a binomial random variable, 123
of moment-generating function for a normal random variable, 123
of multivariate data, 138
of mutually exclusive events, 31
of negative binomial distribution, 59, 110, 118
of normal approximation to binomial, 81
of normal distribution calculations, 76–77, 77*f*
of normal distribution MLE for mean and variance, 165
of normal distribution moment estimators, 162
of normal probability plot, 145, 257*f*
of Ohm's law, 12, 13
of one-sided confidence bound, 176
of Poisson distribution, 230–231, 230*t*, 231*t*
of polymerization reactor, 12
of possible number of sequences, 25
of probabilities of events in a random experiment, 28
of probability and sensitivity of a test, 40
of probability as a ratio of volumes, 108
of probability density function, 69*f*
of probability distribution, 44*f*, 47*f*, 48*f*
of probability distributions for more than one random variable, 96
of probability mass function for discrete uniform random variable, 50, 50*f*
of probability of failure, 39
of probability of product failure, 35
of regression analysis, 301
of sample mean and variance being unbiased, 156–157
of sample mean of data, 127–128, 128*f*
of sample variance, 129, 129*t*
of sampling without replacement, 61
of selected hypergeometric distribution, 50, 60
of series of repeated, random trials, 51
of significance of regression, 323
of simple linear regression model, 286
of standard deviation of lognormal random variable, 92
of standard normal distribution, 76
of stem-and-leaf diagram, 132
of subjective probability or degree of belief, 26
of types of engineering estimation problems, 150
of uniform distribution MLE, 167, 167*f*
of using the addition rule, 30
of variability, 3
of Weibull random variable, 90
of Wilcoxon signed-rank test, 239
organic pollution in water, 54
orthopedic physician office bill number of errors, 105
oxygen purity, 286–287, 286*f*, 287*t*, 298*f*, 298*t*
oxygen purity ANOVA, 292
oxygen purity confidence interval on the mean response, 294, 294*f*
oxygen purity confidence interval on the slope, 293
oxygen purity prediction interval, 296, 296*f*
oxygen purity residuals, 297, 298*f*, 298*t*
oxygen purity tests of coefficients, 287*t*, 290
paint drying time sample size, 250
paint drying time sample size from OC curves, 249
paper tensile strength ANOVA, 356, 358*t*
parallel circuit, 38, 38*f*
particles of contamination on semiconductor wafer, 44
parts from two circuit board suppliers, 61

patient satisfaction in a hospital, factors that affect, 311
patients with illness who improve from medication, 51
percentage of hydrocarbons in condenser of distillation unit, 281–282, 282*f*, 282*t*
percentage of process yield with two variables, 407, 407*f*, 407*t*
pinot noir wine quality, 343–345, 343*f*, 344*t*, 345*t*
pressure of gas in container and temperature and other factors, 281
printed circuit board defects, 230–231, 230*t*, 231*t*
printed circuit board design layout for five components, 26
printed circuit board design layout for four components, 25
printed circuit board mechanical and electrical components, 459, 459*f*, 459*t*
process yield control composite design, 428–431, 429*f*, 429*t*, 430*f*, 430*t*, 431*f*
propellant shear strength, 239
propellant shear strength sign test, 236, 236*t*
proportion of people who like a particular soft drink, 151
proportion of solar radiation absorbed by a material, 92
proportion of voice lines in use, 51
pull strength of a wire bond in semiconductor manufacturing, 315–316, 315*t*, 316*t*
pull-off force for connectors in automobile, 73
pull-off force of engine connectors, 127–128, 128*f*
pull-off force variability, 4, 4*f*, 7
quality of young red wines, 142–144, 142*t*, 143*f*
RAID to increase robustness to disk failure, 43
reaction temperatures in a chemical process, 149
reaction time of a chemical, 71
reducing the drying time of primer paint, 248
resistivity of semiconductor wafers, 11
risk of data loss in a redundant system, 43
rocket propellant burning rate sample size from OC curve, 214
rocket propellant burning rate type II error, 213, 214
sampling with replacement, 36–37, 60
sampling without replacement, 26, 37
semiconductor automobile engine controller and process fallout, 226, 228
semiconductor chip contamination, 39
semiconductor chip contamination and product failure, 35, 35*t*
semiconductor etch variability sample size and two gases, 272
semiconductor laser lifetime of product, 91–92
semiconductor manufacturing quality index at three plants, 139*f*, 140
semiconductor product and variance of three layers, 119
semiconductor wafer and sources of contamination, 29, 30*f*
semiconductor wafer with large particles of contamination, 45
semiconductor wafers, 11–12, 15
semiconductor wire-bonded to a frame, 12, 13*f*
serial number and its digits, 50, 50*f*
series circuit, 38, 38*f*
shear strength for steel plate girders, 265–266, 265*t*
shear strength of spot welds to weld diameter, 299
soft-drink beverage bottler bursting strength of glass bottles, 202
Space Shuttle *Challenger* and failure of O-rings, 281, 307–308, 307*f*, 307*t*, 308*f*
St. John's wort to treat depression, 275
surface finish of metal parts on a lathe, 340–341, 340*t*, 341*t*
surface flaws and defective parts, 32, 32*t*
surface roughness of a part in a metal-cutting operation, 389, 389t, 399–401, 399*t*, 400*t*, 402*f*
synthetic fiber breaking strength in body armor for military/police, 242
tensile adhesion tests on U-700 alloy, 191, 192
tensile strength of component in air frame, 149
tensile strength of different grades of aluminum spars for aircraft wings, 252
tensile strength of paper (psi), 353–354, 353*t*, 354*f*
thermal conductivity of Armco iron, 159
thin film manufacturing process and contamination, 32
time needed for parallel parking for cars with different wheel bases and turning radii, 267–268, 268t
time needed for service of constant-velocity joint in car, 93
time needed to prepare a slide for high-throughput genomics, 88
transportation and trips, 2
trip-generation model, 2
two cameras flash recycle time, 20, 22
two catalysts and effect on mean yield of chemical process, 255–256, 255*t*, 256*f*
uniform current and measure of current, 73, 73*f*
use of probability model, 2
use of regression analysis, 2
use of the multiplication rule, 24, 26
vane opening jet engine part and investment casting process, 446–449, 447*t*, 448*f*, 449*f*, 449*t*, 453*f*
voice communication system and number of phone lines, 19
water contamination with a large particle, 56
Web site design for color, font and position of an image, 24
Web site mobile response time for downloads, 96, 97*t*
windmill to generate electrical power, 303–305, 303*f*, 304*f*, 304*t*, 305*t*
wine quality stepwise regression, 346–347, 347*t*
wire bond pull strength, 13, 13*f*, 14*f*, 301–302, 302*f*
wire bond pull-strength general regression test, 328
wire bond strength ANOVA, 323, 324*t*
wire bond strength coefficient test, 325–326
wire bond strength confidence interval, 330, 332
wire bond strength confidence interval on the mean response, 331
wire bond strength Cook's distances, 336, 337*t*
wire bond strength one-sided coefficient test, 326
wire bond strength residuals, 333, 333*f*
wire bond strength with matrix notation, 318–319, 318*t*, 319*t*, 320*t*
wire flaws in thin copper wire, 63
worn machine tool and defective parts produced, 51
Exhaustive event, 36
Expected mean squares, randomized complete block experiment, 370
Expected value
also called the mean, 47
of a function of a continuous random variable, 72
of a function of a discrete random variable, 49
of a function of two random variables, 110

Expected value (*contd.*)
of mean squares, single-factor experiment, 356
of mean squares, two factors, 384
Experiment. *See also* Study
2^k factorial
binomial, 54
bowl of beads, 435
comparative, 7, 246
complete block, 370
completely randomized, 246
completely randomized single-factor, 353–365, 354*t*
definition of, 376
design of a, 376
designed, 5, 6–7, 7*f*, 284
factorial, 7, 7*f*, 8, 376, 378–381, 379*t*, 380*f*, 381*f*
four-factorial, 8*f*
fractional factorial, 8–9, 9*f*
optimization, 377, 378*f*, 381*f*
process characterization, 376
randomized block, 371
randomized complete block, 370
screening, 353, 413
single-factor, 352
statistical analysis of a two-factor, 382
statistical methods/thinking and engineering, 7
two-factor factorial, 382–388, 382*t*
unbalanced, 359
with a random component, 18
Experimental design, tool for engineers and scientists, 376
Experimental design methods, 352
Exponential distribution, 83–84
maximum likelihood estimator and, 164
used in reliability studies for time to failure of a device, 86
where the name came from, 83, 84*f*
Exponential random variable, 83, 85
first event in a Poisson process and an, 86
is the continuous analog of a geometric random variable, 86
Exponentially weighted moving-average (EWMA), also called geometric average, 462, 463*f*
Extra sum of squares method, 326
Extrapolation
be careful when going beyond original observation to do a large, 284, 332
hidden, 332
in multiple regression, 332
of a regression model, 284
Extreme outlier, 139, 139*f*

F

F distribution, 268–270, 269*f*
finding lower-tail points of the, 270
probability density function of, 269*f*
upper and lower percentage points of the, 269*f*
for general regression tests, 327
F-random variable, 346
F test, for effects, 384
Factor
correction, 62, 81
estimate the magnitude and direction of effect of a, 377
fixed versus random, 365
finite population correction, 62
fixed versus random, 365
noise, 377
nuisance, 368
uncontrollable, 377
variance inflation (VIF), 349
Factor levels, definition of, 7
Factor of interest, 352
Factorial, two-factor, 385*t*
Factorial design, 7*f*, 8*f*
Factorial experiment, 376, 378–381, 379*t*, 380*f*, 381*f*
2^k, 8
advantages of a, 8
definition of, 7, 7*f*, 378
how to calculate the number of runs needed, 8
only way to find interaction between variables, 380
used in industrial research and design, 9
with interaction, 379*f*
with two factors, 379*f*
Failure, label that can be misleading, 51
Failure mechanism, 144
Failure time, Weibull distribution used to model device, 86
Fallout, definition of, 368
False alarm, 452, 460
False null hypothesis, 201
Finite population correction factor, 62
Fisher method, *P*-values and, 242
Fisher, R. A., 242
Fisher's least significant difference (LSD) method, 359–360, 372–373, 373*f*
Fitted regression equation/model, 283, 318
Fitted value, 318, 319*t*
Fixed factor, 365
Fixed significance level, 218
Fixed significance level testing, 203
Fixed-effects model, 355
Fluid science, 3
Forward selection, variation on stepwise regression, 347
Fraction
alternate, 414, 414*f*
principle, 414, 414*f*
Fraction defective, 226, 228
Fraction-defective control chart, 456
Fraction generator, 413
Fractional factorial, main use is in screening experiments, 413
Fractional factorial design, 413
alias pattern of, 417
design resolution for, 417
Fractional factorial experiment
definition of, 8–9, 9*f*
used in industrial research and design, 9
Fractional factorial saturated, 424
Fractional replication, 413, 414
Freedom, degrees of, 130, 183
Frequency, relative, 29
Frequency distribution, 135–136, 135*t*
data divided into class intervals, cells, or bins, 135
visually displayed by a histogram, 136
F-tests, 234
for significance of regression, 350
only allows testing against a two-sided alternative hypothesis, 292
P-values for the, 271–272
Full model, 327
Function. *See* Specific function name
Funnel experiment, equipment and details of Deming's, 10, 10*f*

G

Galton, Sir Francis, 283
Gamma distribution, 145
Gamma distribution moment estimators, 163
Gamma function, equation for, 87
Gamma probability density function, 88*f*
Gamma random variable, 87
calculate the probabilities for a, 88
mean and variance of a, 88
Gas, distribution of molecular velocity at equilibrium in a, 67
Gasoline mileage, variability in, 3
Gauss
developed the normal distribution, 73
Karl, 13
least squares method based on work of, 13
method of least squares created by Karl, 284
Gaussian distribution, also called a normal distribution, 73
General factorial experiments, 388–390, 388*t*
General regression significance test, 326
General regression test
F statistic and, 327
hypotheses for, 326
Generalized interaction, 412
Generating function, moment-, 121–124
Generator, of a fraction, 413
Generators, design, 419

Geometric average (geometric mean), 462
also called exponentially weighted moving-average, 462
Geometric distribution, 55–57, 56*f*
quality of lack of memory and, 57
Geometric notation, definition of, 390
Geometric random variable, 55, 110, 460
Bernoulli trial and, 55, 56*f*
equation for, 56
how to obtain the variance of a, 56
mean of, 56
probability mass function for, 55
Goodness-of-fit test, 229–232
Goodness-of-fit test statistic, 230
Grand mean, 444
Graphical comparison, of means, 360
Graphical display
for standard normal distribution, 77*f*
of partitioning an event, 36*f*

H

H matrix (hat matrix), 334–335
"Hat," 13
Hat matrix (H matrix), 334–335
diagonal elements of a, 335
Hidden extrapolation, 332
Histogram
advantages and disadvantages of, 136
an SPC problem-solving tool, 436
approximates a probability density function, 68*f*
definition of, 68
for symmetrical and skewed distributions, 138*f*
how to construct a, 136
mercury concentration of largemouth bass, 178*f*
symmetrical bell-shaped distribution of measurements, 136, 136*f*
used to assess process capability, 452
visual display of the frequency distribution, 136
Homogeneity test, 234
Homogeneous population, 234
Hunter, J. Stuart, 140
Hypercube, 8
Hypergeometric distribution, 59–63, 61*f*
binomial distribution and, 81
compared to binomial distribution, 62, 62*f*, 62*t*
for selected variables of parameters, 61*f*
Hypergeometric random variable, 60, 62
mean and variance of, 61
Hyperplane, 312
Hypothesis
alternative, 195, 206
an equality claim and null, 195
binomial proportion, and testing, 226
definition of, 7, 195
false null, 201
for ANOVA test, 322
for general regression test, 326
lower-tailed test and, 209
never about the sample, 195
null, 206
one-sided alternative, 195
one-sided and two-sided, 202–203
statistical, 194–196
test of, 196
two-sided alternative, 195, 208, 209*f*
two-tailed test and, 208–209
Hypothesis tests
in multiple linear regression, 322–329
in simple linear regression, 288–292
on the difference in means, variances unknown, 253–259
on the equity of two variances, 270–272, 271*f*
on the mean, 208–211, 215–220
on the variance, 222–223, 222*f*
P-values in, 203–206, 204*f*, 205*f*
Hypothesis testing, 7, 149, 188, 194–207, 196*f*, 197*t*, 198*f*
confidence intervals and relationship to, 206
objective of, 195
on difference in means, variances known, 247–248
Hypothesis testing methodology, steps in applying, 206–207
Hypothetical population, 128

I

Ideal gas law, 5
Independence, 36–39, 104–106
bivariate normal random variable and zero correlation implies, 117
conditional probability and, 36
multiple random variables and the concept of, 109–110
of multiple events, 37
of two events, 37
Independence relationship, probability model and, 37
Independent normal random variables, linear function of, 120
Independent populations, 246*f*
Independent random variables, 105–106
probabilities for multiple variables are easier to compute, 106
Indicator variable
categorical regressors and, 339–341
for different levels of qualitative variables, 339
only one in a simple linear regression model
Individuals control chart, 450
Inference, 206
assumptions for two-sample, 246–247
on the difference in means of two normal distributions, variances known, 245–261
on the variances of two normal distributions, 268–273
on two population proportions, 273–278
statistical, 15
Inference procedure
for a single sample, 229
summary table and road map for two samples, 278–279, 279*t*
Influential observations, 335
Influential points, 335
Information, data versus useful, 6
Input, controllable and uncontrollable, 18
Integer values, 46
Interaction
AD (gap-power), 403*f*
between factors, 8
between variables, 380
examples of, 312
factorial experiment with, 379*f*
generalized, 412
Interaction plot, two-factor, 380
Intercept, 312
and slope, 283
confidence intervals on, 292–293
Interquartile range, 135, 139, 139*f*
Intersection
definition and notation for, 21
type of set operation, 21
Intersection of events, 29, 34–36
Interval
confidence, 171–172
decision, 468–470
prediction, 172, 189–191
tolerance, 172, 191–192
Interval estimates, types of, 171, 172
Intrinsically linear model, 303
Invariance property, 166
Iteration, or cycle, 3

J

Jacobian, transformation and the, 121
Joint distribution, mean and variance from a, 101, 108
Joint events
calculating the probability of, 29
types of, 29
Joint probability density function, 107
definition of, 97–98, 98*f*
Joint probability distribution, 95–125
defines simultaneous behavior of two random variables, 96
for more than two random variables, 107–110
for two random variables, 96–101
sign of covariance and, 111*f*

Joint probability mass function, 97
describes the joint probability distribution of two discrete random variables, 97
Juran, Joseph M., 436

K

Karlsruhe procedure, 265–266

L

Lack of memory property, 57, 59, 85–86
exponential distribution and, 86*f*
exponential random variable and, 85
probability distribution of random variable and, 57
Lack-of-fit testing, use as final check on model-building, 348
Large sample approximation, 240, 263
Large-sample test, 215, 225–227, 248
on a proportion, 225–227
on difference in population proportions, 274–275
Law
DeMorgan's, 23
examples of physical, 5
ideal gas, 5
movement from a specific case to a, 5
Newton's, 18
Ohm's, 5, 12, 19
set operations and distributive, 23
Leaf, definition of, 132
Least significant difference (LSD) method/test, Fisher's, 372–373
for multiple comparisons, 360
Least squares
estimate of β, 317
to estimate regression coefficients, 284
used to estimate regression coefficients in multiple regression model, 314
Least squares estimates, 285, 314
Least squares estimation, of parameters, 314–316
Least squares estimators, properties of, 288, 321
Least squares fitted model, 394
Least squares line, called regression line by Galton, 283
Least squares method
based on work of Karl Gauss, 13
used to estimate parameters in regression models, 13
Least squares normal equations, 284–285
in matrix form, 317
Lehigh procedure, 265–266
Length
average run (ARL), 460–461, 461*t*
example of continuous random variable, 41
Poisson distribution and, 64
Level of significance, 203, 247, 250
Levels, of a factor, 352
Likelihood, method of maximum, 300, 307
Likelihood function, 163–167
Limit
3-sigma control, 460
average run (ARL), 460–461, 461*t*
natural tolerance, 453
warning, 443
Line, regression, 283, 283*f*, 284, 285*f*
Linear association, 144
Linear function
mean and variance of a, 118
of independent normal random variables, 120
of random variables, 117–119
of unknown parameters, 312
Linear model
intrinsically, 303
probabilistic, 281
Linear permutations, 24
Linear regression, works best when response variable is quantitative, 305
Linear regression model, definition of, 313–314, 313*f*
Linear response function, 306, 306*f*
Listing, the elements in a set, 25
Location parameter, 152
Log
condenser temperature controller, 5
reboil temperature, 5
record over time, 5
Log likelihood, 164–165, 165*f*
Logical point estimator, 247
Logistic regression, 305–310
Logit response function, 306
Lognormal distribution, 91
central limit theorem and, 153*f*
Lognormal probability density function, 91*f*
Lognormal random variable, 91
Lot quality, probability model and, 15
Lower control limit (LCL), on a control chart, 437, 443
LWBS, leave without being seen, 22
LWOT, leave without receiving treatment, 2

M

"Magnificent seven" problem-solving tools, 476
Main effect, definition of, 379
Mann-Whitney test, another name for Wilcoxon rank-sum test, 261
Marginal distribution, of bivariate normal random variables, 116
Marginal probability density function, 100, 106, 108
Marginal probability distribution, 109, 109*f*
definition of, 99, 100*t*
Mathematical model
also called an abstraction, 18
of a physical system, 18, 19*f*
Matrix
covariance, 321
design, 396
of scatter diagrams, 143, 143*f*
Matrix approach, to multiple linear regression, 316–321
Matrix notation, 316
Maximum likelihood estimation, complications of using, 166–167
Maximum likelihood estimator (MLE), 163, 165–166
Maxwell, James, 67
Mean
also called expected value, 47
definition and equation for, 47
hypothesis tests on the, 208–211, 215–220
large-sample confidence interval for the, 177
of a linear function, 118
of geometric random variable, 56
overall, 355
population, 130
sample, 151
sampling distribution of the, 151–152
trimmed, 157
Mean and variance
conditional, 104
from a joint distribution, 101, 108
normal distribution MLEs for, 165
normal probability density functions and, 74*f*
of a beta distribution, 93
of a binomial random variable, 54–55
of a continuous random variable, 71
of a continuous uniform random variable, 73
of a discrete random variable, 47–49
of a discrete uniform random variable, 50–51
of a gamma random variable, 88
of a geometric random variable, 57
of a hypergeometric random variable, 61
of a negative binomial random variable, 59
of a Poisson random variable, 65
of a random variable with an exponential distribution, 83
of a Weibull distribution, 90
of an average, 119
of an Erlang random variable, 118
of an χ^2 distribution, 183
of maximum likelihood estimators, 166
Mean of a normal distribution, variance known, tests on the, 208–215
Mean response, confidence interval on the, 293–294, 330–331

Mean square
or error (SS_{Error}), 356
for treatments ($MS_{\text{Treatments}}$), 356
Mean squared error, of an estimator, 160–161
Mean squares
how to compute, 291
random effects and expected values of, 366
Means
graphical comparison of, 360
tests on individual, 386
Mechanical science, 3
Mechanistic model, 12–13
definition of and equation for, 12
Median
sample, 133
sign test used to test hypothesis about the, 235
Meta-analysis, combining results from several studies, 242
Method
engineering, 2–5, 3*f*
least squares, 13
scientific, 2
Method of maximum likelihood, 163–167
and point estimators,161
Method of moments
and point estimators, 161
population moments and sample moments, 162–163
Method of steepest ascent, 426, 427*f*
Method of steepest descent, minimization and, 426
Minimax criterion, definition of, 474
Minimin criterion, definition of, 474
Minimization, method of steepest descent and, 426
Minimum variance unbiased estimator (MVUE), 158
Mirroring, definition of, 43
M_n, average molecular weight of a polymer, 12
Mode, peak of density, 93
Model
binomial, 79
collect data to test a tentative, 2, 3*f*
continuous iteration between physical system and, 19*f*
decision, 473
empirical, 13, 281–284
first-order, 425
fitted first-order, 426
fitted regression, 283
fixed-effects, 355, 385*t*
full, 327
in fitting polynomials use the lowest-degree, 338
intrinsically linear, 303
least squares fitted, 394
limitations or assumptions of a, 2, 3*f*
linear regression, 313–314, 313*f*
manipulate the, 2, 3*f*
mathematical, 18
mechanistic, equation for, 12–13
multiple regression, 314
nonlinear, 303
polynomial regression, 337–339
probabilistic linear, 281
probability, 15
quadratic, 430*t*
random-effects, 365–368
reduced, 327
regression, 2, 13, 285*f*
second-order, 313–314, 338, 338*t*, 426
simple linear regression, 282, 286, 296
that includes variation, 18
that incorporates uncontrollable inputs, 18
trip-generation, 2
true regression, 283
use observed data to refine a, 2, 3*f*
variance, 365
Model adequacy checking, 333–337, 386–387
Model and residual analysis, of a 2^2 factorial design, 401
Model and residual analysis, of a 2^k factorial design, 393
Model matrix, also called the X matrix, 317
Model parameter
estimating the, 296
testing a statistical hypothesis and, 288
Moment, population, 162
Moment about the origin, definition of, 122
Moment estimator, exponential distribution, 162
Moment-generating function, 121–124
definition of, 122
for a binomial random variable, 123
for a normal random variable, 123
properties of the, 124
Montgomery, Peck, and Vining, 5, 297, 303, 307, 342, 350
Montgomery, Peck, Vining, and Myers, 336
Moving average, definition of, 462
Moving range, 450
Multicollinearity
strong dependency among regressor variables and, 349
why it occurs, 349–350
Multinomial probability distribution, 113–114
Multinomial random experiment, 114
Multiple comparisons, following ANOVA, 359–361
Multiple comparison method, examples of, 372, 359
Multiple linear regression, 310–350
confidence intervals in, 329–331
examples of the use of, 311
matrix approach to, 316–321
Multiple linear regression model
contains more than one regressor variable, 311, 314*t*
with k regressor variables, 312
Multiple regression model, fitting of a, 316
Multiple regression modeling, aspects of, 337–350
Multiplication rule
a counting technique, 24
conditional probability and the, 34
equation for, 34
to determine number of outcomes in a sample space or event, 24–26
Multivariate data, 138
Mutually exclusive, when two events are, 30
Mutually exclusive events, 31
definition of, 23
Mutually exclusive subsets, partition of an event into, 31, 36, 36*f*

N

Natural tolerance limits, 453
Negative binomial distribution, 57–59, 110
Bernoulli trial and, 57–59
for select values of parameters, 58*f*
Negative binomial random variable, 58–59, 59*f*
as opposite of binomial random variable, 59
as sum of geometric random variables, 59*f*
mean and variance of, 59
Newton's law, as model of physical universe, 18
Noise, example of an uncontrollable input, 18, 19*f*
Noise factors, are uncontrollable factors, 377
Noncentral *F*-distribution, 363
Nonconforming parts, 59–60
Nondefective parts, and defective parts, 26
Noninteger values, 46
Nonlinear relationship, between random variables, 111
Nonlinearity, can be visually determined from a scatter diagram, 303
Nonnegative integer, 63
Nonparametric approach, to paired comparisons, 268
Nonparametric procedures, 234–240
less efficient than parametric procedure, 234
Nonparametric test
difference in two means and a, 261–264
for the difference in two means, 261–264
Nonrandom shift, 10
Nonzero probability, 46

Normal approximation, 79–80, 80*f*
binomial distribution and, 80–81, 80*f*
for a binomial proportion, 185
for the sign test statistic, 237
for Wilcoxon rank-sum test statistic, 263
for Wilcoxon signed-rank statistic, 240
Poisson distribution and, 82
to the binomial distribution, 79
to the Poisson distribution, 79
Normal distribution, 73–79, 75*f*, 77*f*
also called a Gaussian distribution, 73
can be used to approximate probabilities of a Poisson random variable, 81
continuous probability distribution and, 67
developed by Gauss, 73
reproductive property and, 119
tests on the variance of a, 223
tests on variance and standard deviation of a, 222–224
used to approximate binomial probabilities, 79
Normal equation, matrix form of is identical to scalar form, 317
Normal probability density function, 75
definition and formula for, 74
mean and variance for, 74*f*
Normal probability plot, 146–147, 147*f*, 182*f*
central limit theorem and, 154*f*
mercury contamination of largemouth bass, 178*f*
of effects, 417
of residuals, 296, 297*f*, 333*f*
of the coefficient of restitution, 219*f*
Normal random variable, 74
moment-generating function for a, 123
most-used model for a continuous measurement, 73
standardizing a, 77, 78*f*
Normality, *t*-test and, 219–220
Normality assumption, 361
Normally independently distributed (NID), 289
Notation, for totals and means, 383
NP chart, computer produces a, 458
Nuisance factor
blocking out a, 368
definition of, 368
Null distribution, 216
Null hypothesis, 195, 206, 248, 254, 255, 258, 262, 264, 265, 270, 271, 274, 275, 278
for a paired *t*-test, 264
part of procedure for hypothesis tests, 206
sign test and the, 268
Null set, 22

O

Objective probabilities, 167
Observation, of processes over time, 9–12
Observational study, 246
advantages and disadvantages of an, 6
definition and example of, 5–6
Observed significance level, 203
Odds, 307
Odds ratio, 307–308
Ohm's law, 5, 12, 19
One-factor-at-a-time approach 380–381, 381*f*
One-half fraction, test that uses only one-half of possible factor combinations, 9
One observation per cell, 387–388
Operating characteristic (OC) curve
definition of, 363, 364*f*
using, 213–214, 249
Operations, on sets, 19. *See also* Set operations
Optimization, definition of, 353
Optimization experiment
and one-factor-at-a-time approach, 381*f*
definition of, 377, 378*f*
Optimize the response, 425
Ordered stem-and-leaf diagram, 133
Orthogonal
columns in a table, 398
columns of a model matrix, 349
Orthogonal design, 401
Outcomes
continuous sample space and interval of real numbers, 20
determining the number in a sample space or event, 23, 24
discrete sample space and countable set of, 20
equally like, 27
random experiment and set of all possible, 20
to analyze a random experiment must understand possible, 19
unbounded number of, 45
Outlier, 139, 139*f*, 297, 335
definition of, 6
Output, combination of uncontrollable input and controllable input, 18
Overall mean, 355
Overcontrol, definition of, 10, 10*f*
Overfitting, definition of, 324–325
Ozone, definition of, 96

P

P chart
also called control chart for fractional nonconforming, 456
center line and upper and lower control limits for a, 457
control chart proportions, 456–458
Paired *t*-test, 264–265
compared to unpaired, 266–267
for Karlsruhe-Lehigh procedure, 256
null hypothesis and test statistic for a, 264
Parameter (θ), 149
Bayesian estimation of, 167–169
confidence interval on, 293
estimation of, 130
linear function of unknown, 312
location, 152
scale and shape, 87, 152
Parameter estimation, 149
Parameter of interest, 206, 248, 255, 258, 262, 265, 271, 275
part of the procedure for hypothesis tests, 206
Parameters least squares estimation of, 314–316
Parametric methods, 234
Parametric procedure, more efficient than nonparametric procedure, 234
Parametric technique, to transform original data, 234
Pareto chart, 476
Pareto chart,
an SPC problem-solving tool, 436
definition of, 138
Pareto diagram, an SPC problem-solving tool, 471, 471*f*
Parity data, definition of, 43
Partial *F*-test, 327
Partial regression coefficients, 312
Partitioning, of an event into mutually exclusive subsets, 36, 36*f*
Parts per million (PPM), 455, 455*t*
Path, of functional devices, 38
Pattern recognition, on a control chart, 443
PCR_k, equation for, 454
Pearson correlation coefficient, 143
Percentage points
F distribution and upper and lower, 269*f*
of a *t* distribution, 180, 180*f*
of an χ^2 distribution, 183
Percentile, 134
Permutation
also known as linear permutations, 24
definition of, 24
how to construct a, 24
of similar objects and equation for, 25
of subsets and equation for, 24
Physical laws, examples of, 5
Physical system
can be modeled by continuous random variables, 67
can be modeled by random experiments and variables, 43
continuous iteration between model and, 19*f*

engineering uses density function to describe, 67
failure due to wear or time or external shocks, 89
mathematical model of a, 18
Piegorsch and Bailer, 242
Plot
box, 139, 139*f*
box-and-whisker, 139, 139*f*, 182*f*
contour, 115, 115*f*
cumulative distribution, 137*f*
cumulative frequency, 137
digidot, 140–141, 141*f*
normal probability, 146–147, 147*f*, 182*f*
of pull strength, 14*f*
of residuals, 334*f*
probability, 144–147, 145*f*, 145*t*
residual, 297, 297*f*, 304
residuals and normal probability, 296, 297*f*
three-dimensional, 13*f*
three-dimensional surface, 380, 381*f*
time-oriented, 9
time sequence, 140–141
time series, 9, 9*f*, 141*f*
two-factor interaction, 380
PM10, particulate matter in the air, 96
Point estimate, 149, 150
Point estimation, 149–150
Point estimation
general concepts of, 156–161
points of, 161–169
Point estimation of parameters, and sample distributions, 148–169
Point estimator, 150, 295
how to obtain good, 161
method of maximum likelihood, 161
method of moments and, 161
variance of a, 157–158
Poisson, Siméon-Denis, 64
Poisson distribution, 64–65, 64*f*, 230
definition of a random variable, 83
normal approximation and, 79, 82
why it was developed, 81
Poisson probabilities, can be approximated with probabilities based on normal distribution, 82
Poisson process, 83, 85, 86, 88
types of random experiments using the, 63
Poisson random variable, 63
distribution of a sum of, 124
mean and variance of, 65
Polymer, average molecular weight of a (M_n), 12, 13
Polymerization reactor, 12
Polynomial regression model, 337–339
Polynomials, use the lowest-degree model in fitting, 338
Pooled estimator, 253–254
Pooled *t*-test, 254
Pooling, of the data, 341, 402
Population
also known as conceptual or hypothetical population, 128
homogeneous, 234
independent, 246*f*
sample and relationship to a, 131, 131*f*
sample selected from a larger, 128
sample versus an entire, 5, 5*f*, 12*f*
true value of the, 130
Population mean, 130
Population moments, 162
Population proportion
approximate sample size and two-sided test on difference in, 276
approximate type II error and one-sided test on difference in, 276
approximate type II error and two-sided test on difference in, 276
confidence interval on difference in, 277–278
inference on two, 273–278
large-sample confidence interval for a, 185–188
large-sample tests on different, 274–275
large-scale confident interval for a, 185
test statistic for difference of two, 274
tests on, 225–229
Population standard deviation, 130
Population variance, 130
Positive quadrant of the plane, 20
Power, of a statistical test, 201
Practical significance, versus statistical significance, 207, 207*t*
Precision of estimation
confidence level and, 174–175, 175*f*
for regression coefficients, 321
Prediction, of new/future observations, 295–296, 331–332
Prediction error sum of squares (PRESS) statistic, 343–345
Prediction interval
definition of, 172, 295
for future observation, 189–191
on a future observation, 295, 331
purpose of, 172
Predictor variable, 281, 284
PRESS statistic, regression models and the, 343
Pressure, example of continuous random variable, 41
Principal block, 409
Principal fraction, 414, 414*f*
Prior distribution, also called probability distribution, 167
Probability, 17–41
as a ratio of volumes, 108
based on the conceptual model of repeated replications, 26
binomial, 79
conditional, 31–34, 32*f*
definition of, 26, 27–28, 27*f*
degree of belief or subjective, 26
for a discrete sample space, 26
interpretations and axioms of, 26–29
nonzero, 46
objective, 167
of a type I error, 197, 201
of a type II error for a two-sided test on the mean, variance known, 212
of a union, 30–31
of failure, 39
Poisson, 82
positive 106
relative frequency interpretation of, 27
rule of total, 35, 36
standardizing to calculate a, 78
subjective, 26, 167
type II error (β), 198–201, 199*f*, 200*f*, 211–212, 212*f*
unconditional, 37
what it is determined by, 28
Probability density factor, of a continuous random variable, 70
Probability density function, 67–69, 68*f*, 83, 87, 90, 91*f*
characteristic symmetric bell-shaped curve, 74
continuous random variable with, 71,72
definition of, 68
how to determine it from a cumulative distribution function, 70
of a bivariate normal distribution, 115
of *F* distributions, 269*f*
of *t* distributions, 180*f*
on a histogram, 68*f*
used to describe probability distribution of continuous random variable, 67
Probability distribution, 44, 67, 67*f*
also called prior distribution, 167
definition of, 43
for a random variable, 47
for bits in error, 44*f*
loading with mean and balance points, 47*f*
of the random variable, 43
Probability mass function, 46, 52, 97, 100, 114, 120
and cumulative distribution function, 46
definition and equation for, 44
for geometric random variables, 55
Probability model
and statistical inference, 15
decreases number of samples needed, 15
definition of, 15
Probability of an event, ratio of number of outcomes in event to that of sample space, 28

Probability plot, 144–147, 145*f*, 145*t*, 146*f*
how to construct a, 145, 146*f*
Probability plotting, 230
Probability space, 110
Probabilities
associated with a normal distribution, 75*f*
conditions for approximating hypergeometric and binomial, 82*f*
cumulative Poisson, 88
for exponential random variables, used in staffing and customer service, 84
properly combining given, 18
sum of all outcomes of an experiment adds up to 1, 27–28, 27*f*
Problem. *See also* Example
engineering method and description of the, 2, 3*f*
factors that affect a, 2, 3*f*
single-sample hypothesis-testing, 7
two-sample hypothesis-testing, 7
Problem-solving tools, types of SPC, 436
Process
6-sigma, 455, 456*f*
historical use of word, 64
Poisson, 83, 85, 86, 88
Process capability, 439
definition of, 452–456
Process capability ratio (PCR), 452–454, 453*f*, 455*t*
Process characterization experiment, 376
Process fallout, 453*f*
Process mean shift, 11*f*
Process variability, control chart is primary tool to understand, 435
Projection property, used in factor screening, 417
Properties, of the least squares estimators, 288
Properties of summation, 47
Property, lack of memory, 57, 59
Proportion
large-sample tests on a, 225–227
sample, 151
tests on population, 225–229
Prototype, 4
Pull strength
definition of, 13
wire bond and data for, 13*f*
Pull-off force measurement/variability, 4, 4*f*, 7
P-value, 203–205, 204*f*, 205*f*, 208
for the *F*-test, 271–272
calculated by statistics software, 217
calculated for a *t*-test, 216*f*, 217, 217*f*
combining, 242
equals the observed significance level, 203, 204*f*
Fisher method to combine all individual, 242
for a *z*-test, 209*f*
in hypothesis tests, 203–206, 204*f*, 205*f*

Q

Quadratic model, analysis of variance for the, 430*t*
Qualitative response variable, 305
Quality, means fitness for use, 435
Quality control, statistical, 434
Quality improvement
elimination of defects and waste, 435
reduction of variability, 435
statistics and, 435–436
Quality of conformance, 435
Quality of design, 435
Quantitative response variable, 305
Quantitative variable, examples of, 339
Quartile, definition of, 134

R

R chart, estimator of mean of variance from, 444
R control chart, computer construction of, 449
R^2, adjusted statistic, 324
R^2, coefficient of determination, 298–299
RAID, redundant array of independent disks, 43
5 design and 6 design, 43
Raleigh distribution, 89
Randomized block/complete block experiment, 370, 371
Random component, of an experiment, 18
Random disturbance, 4
result of adjustments to, 10, 10*f*
Random-effects model, 365–368
Random error, 282, 284, 311, 317
Random factor, fixed versus, 365
Random order, 246, 248
Random sample
and conditional probability, 34
definition of, 34, 151
purpose in taking a, 151
Random variable, 40–41. *See also* Variable
chi-square, 183
definition and equation for, 4, 40
definition of continuous, 41
definition of discrete, 41
independent, 98, 105, 110, 113
notation of before and after an experiment, 40
probability distribution of the, 43
Randomization, 6
runs and the role of, 353
Randomized complete block design (RCBD), 368–374, 369*f*, 369*t*, 371*t*
residuals for, 373, 373*t*, 374*f*
Randomly, definition of, 27
Randomness, and selection of a sample, 15
Range, interquartile, 135, 139, 139*f*
Range (*R*) control chart, to control process variability, 444–445
Ranks, data in the form of, 234
Rare event, 155
Rate, false-negative or false-positive, 18
Rate, reflux, 5, 8*f*, 9*f*
Rational subgroups, 441–442
Shewhart concept of, 441
Reactor, polymerization, 12
Reasoning
statistical inference as a type of, 5*f*, 5
types of, 5*f*
Rectangle, area and height, 136
Rectangular range, 106
Reduced model, 327
Redundant array of independent disks (RAID), 43
Reference distribution, 209, 216
for a test statistic, 247
Reference value, 468
Reflux rate, 5, 8*f*, 9*f*
Region
acceptance, 196
critical, 196, 198*f*, 206, 209, 210, 216, 218, 222*f*, 226, 240
rejection, 210
Regression
abuses and misuses of, 284
curvilinear, 357
logistic, 305–310
multiple linear, 310–350
on transformed variables, 303–305
significance of, 289–290, 290*f*
simple linear, 284–289
stepwise, 345–346
Regression analysis, 301, 328*t*
definition of, 281
statistical technique of, 2
term first used by Galton, 283
Regression coefficient, 281, 312
confidence intervals and individual, 329–330
method of least squares and, 284
precision of estimation for, 321
subsets and test on individual, 325–329
Regression equations, 342
Regression line, 283, 283*f*, 284, 285*f*
fitted or estimated, 285
Regression model, 13, 285*f*
2^k factorial design and, 393–394
adequacy of the, 296–299
assumptions needed to fit, 296
can be used to predict future observations, 331
coefficient of determination used to judge adequacy of, 298
most widely used statistical technique, 2

use least squares method to estimate parameters in a, 13
Regression modeling, computer software programs widely used in, 287
Regression of sum of squares, 323
Regression plane, 312*f*
Regression relationships, only valid for regressor variables within range of original data, 284
Regression sum of squares, 291
computer software calls it "model" sum of squares, 292
Regression variables, model building and selection of, 341–349
Regressor, 281
candidate, 347
Regressor variables, 312, 349
Rejection region, 210
Relationship
cause-and-effect, 6
nonlinear between random variables, 111
Relative efficiency, 161
Relative frequency, 27*f*, 29
Relative frequency distribution, 135
Reliability, definition of, 171
Reliability study, exponential distribution and, 86
Replacement, sampling with or without, 26
Replicate
and only one observation per cell, 387
or observation, 353
Replication
fractional, 413, 414
of a random experiment, 26
Reproductive property, of the normal distribution, 119
Residual, 362*f*, 362*t*
definition of, 286, 361
from a regression model, 296
how to standardize, 297
normal probability plot of, 296, 297*f*, 333*f*
partial, 333
standardized, 334
studentized, 334, 336
Residual analysis, 296–297, 333–335, 333*f*, 334*f*
and model, 393
model checking and, 361–363, 362*t*, 373–374, 373*t*
use as final check on model-building, 348
Residual plots, 297, 297*f*, 304
Residual sum of squares, 292, 320
Resolution designs, III, IV, and V, 418
Response surface, 425, 425*f*
analysis of a second-order, 428
first-order, 426
Response surface methodology (RSM)
and designs, 425–432
is a sequential procedure, 426
use of, 425
Response surface plot, for first-order model, 427*f*
Response variable, 284, 305
Retrospective study
acetone-butyl alcohol distillation column example of a, 5–6
problems with historical data derived from a, 6
use of historical data in a, 5
Ridge regression
alternative to ordinary least squares, 350
useful in combating multicollinearity, 350
Romig, Harry G., 436
Rotable, axial spacing and central composite design, 432, 432*f*
Row average, 386
Rule
addition, 30
equation for multiplication, 34
multiplication, 24
set operations and the distributed, 30
total probability, 35, 36
Run, definition of a, 442
Run down, definition of, 442
Run rules, also known as Western Electric zone rules, 443
Run up, definition of, 442

S

S chart, estimator of mean of variance from, 446
s^2, how to compute, 129–130
Sample
bootstrap, 160
data is usually a, 5
determine risk of decision based on a, 15
population and relationship to a, 131, 131*f*
random, 151
randomness and selection of a, 15
selected from a larger population, 128
test, 5
versus entire population, 5, 5*f*, 12*f*
Sample correlation coefficient, 300
definition of, 143
Sample event, 18–23
Sample mean, 151
approximate sampling distribution and, 156
average value of all observations of a data set, 128
definition and equation for, 127
Sample median, definition of, 133
Sample mode, definition of, 133
Sample moments, 162
Sample proportion, 151
Sample range, 131
Sample size
choice of, 175, 175*f*, 187–188
determining the, 363–365
for a confidence interval on the difference in means, variances known, 252
for a one-sided test on the difference in means, variances known, 250
for a specified error on a binomial population, 187
for a two-sided test on the difference in means, variances known, 250
for specified error on the mean, variance known, 175
type II error and choice of, 220–222, 244, 227–229, 249–250, 259–260, 272, 276
Sample size formula, 249–250
for a one-sided test on the mean, variance known, 213
for a two-sided test on the mean, variance known, 212
Sample space (*S*), 18–23
as description of possible outcomes, 40
can be described graphically as tree diagram, 21, 21*f*
defined by objectives of the analysis, 20
definition of continuous and discrete, 20
discrete, 27–28
number of branches in last level of tree diagram equals size of, 23
related outcomes described as subsets of, 21
types of, 20
Sample standard deviation (*S*), 151
positive square root of the sample variance, 128
Sample variance (S^2), 151
how it measures variability, 128–129, 129*f*
standard deviation and, 128, 129*t*
Sampling, with or without replacement, 26
Sampling distribution
central limit theorem and, 150–151
is the probability distribution of a statistic, 149
of the mean, 151–152
Sampling error, definition of, 5
Saturated fractional factorial, definition of, 424
Scale parameter, 152
Scatter
on a dot diagram, 4
or variability of data, 4
Scatter diagram, 142–144, 143*f*
an SPC problem-solving tool, 436
definition of, 281
matrix of, 143, 143*f*
Scatter plot
matrix of, 343*f*
matrix of computer-generated two-dimensional, 316*f*

Science
  electrical, 3
  fluid, 3
  mechanical, 3
  thermal, 3
Sciences, use of engineering method in different, 3
Scientific method, 2
Scientific objectivity, comes from using statistical methods, 353
Screening experiment, 353
  definition of, 413
  use fractional factorials, 413
  use resolution designs III and IV, 418
Sensitivity
  definition of, 40
  of a medical test, 40
Sequences, of fractional factorial design, 415
Sequencing, hospital operating schedule and, 25
Set
  mutually exclusive and exhaustive, 36, 36*f*
  null, 22
  point, 6
Set operations
  distributed rule and, 30
  distributive law for, 23
  types of, 21
Sets, 19
  diagram portrays relationship between, 22
Shewhart, Walter A., 436
Shewhart control chart, 438, 460
disadvantage of the, 462
Shift, nonrandom, 10
Sign test, 235–238
  calculation for the, 237–238, 238*f*
  ties in the, 237
  type II error and the, 237–238
  used to test hypothesis about the median, 235
Sign test statistic, normal approximation for the, 237
Signed-rank test, Wilcoxon, 239
Significance level, 203, 247, 250
  also called type I error probability, 197
Significance of regression, 289–290, 290*f*
  analysis of variance to test, 291–292
  for the second-order model, 338*t*
  test for, 322–325, 323*t*, 324*t*
Simple linear regression, 284–287
  correlation and, 280–309
  hypothesis tests in, 288–292
Simple linear regression model, 282, 286, 296
Single-factor experiments, design and analysis of, 351–374
Single replicate, of the 2^k design, 402–405
Skewed data, description of, 138
Slope
  confidence intervals on, 292–293
  intercept and, 283
  test statistic for, 289
Source, of variability, 3
Space, sample, 18–23
Sparsity of effects principle, 402
SPC, statistical control process, 12
Sputtering tool, 29
Stable process, 11
  upper and lower control limits with a, 11
Standard error, of the point estimator, 185
Standard deviation (S), 71
  and discrete random variable, 47
  continuous random variable and, 71
  equation for, 47
  population, 130
  sample variance and, 128, 129*t*
  when units of random variable are same as units of, 48
Standard deviation chart, to control process variability, 444–446
Standard error
  bootstrap, 159–160
  estimated, 159, 288
  of a coefficient, 394
  of an estimator, 158
  of the point estimator, 185
Standard normal distribution, 76
  is reference distribution for a test statistic, 247
Standard normal probability density function, 76*f*
Standard normal random variable (Z), 75, 77*f*, 91
  how to calculate probabilities for, 77
Standardize, residuals and how to, 297
Standardized residual, 334
Standardizing, definition of, 77
Statistic
  adjusted R^2, 324
  C_p, 342
  definition of a, 149, 151
  prediction error sum of squares (PRESS), 343–345
Statistical analysis
  first step is plotting the data, 127
  of a two-factor experiment, 382–386
Statistical control, definition of, 437
Statistical hypothesis, 194–196
Statistical hypothesis testing, 149, 196–202
Statistical inference, 15
  and probability model, 15
  definition of, 5, 5*f*
  for two samples, 244–279
  major areas of, 149
  use of, 149
Statistical intervals, for a single sample, 170–192
Statistical methods, engineering experiments and, 7
Statistical process control (SPC), 12
  branch of statistics that uses control charts, 12
  definition of, 436
  elements of a company-wide program, 477
  goal to eliminate variability in the process, 437
  implementing a, 476–478
  objective of, 476–478
  types of problem-solving tools, 436
Statistical programs, 127
Statistical quality control, 434–479
  pioneers in the field of, 436
Statistical significance, versus practical significance, 207, 207*t*
Statistical software packages, 127
Statistical thinking, 2–5
  engineering experiments and, 7
Statistically designed experiments, advantage of, 353
Statistics
  definition of, 2, 3
  descriptive, 126–147
  quality improvement and, 435–436
  tell which sources of variability are most important, 3
  the science of data, 127
Steepest ascent, method of, 426, 427*f*
Steepest descent, minimization and method of, 426
Stem, definition of, 132
Stem-and-leaf diagram, 131–132, 132*t*, 133*f*
  can display, percentiles, quartiles, and the median, 133
  computer generated, 133*f*
  how to construct a, 132
Stem-and-leaf display, 134*f*
Steps, in the engineering method, 2–3
Stepwise regression, most widely used variable selection technique, 345–346
Striping, definition of, 43
Studentized residual, 334, 336
Study. *See also* Experiment
  analytic versus enumerative, 11, 12*f*
  analytic, 11
  enumerative, 11
  observational, 5–6
  retrospective, 5
Study objective, 6
Subgroups, rational, 441–442
Subinterval, and Bernoulli trial, 63
Subjective probabilities, 167
  also known as degree of belief, 26
Subset
  factor combination tested is only a, 8
  statistical inference and a data, 15
  when measurements are based on a data, 8, 15
Subsets
  combinations and, 25

partition of an event into mutually exclusive, 31
related outcomes of random experiment described as, 21
Success, label that can be misleading, 51
Sum of squares
error, regression, and total, 291
for an effect, 392
partitioning of variability of data into, 356
Sum of squares identity, 355
Summation, properties of, 47
Surface flaws
defective parts and, 32*f*, 32*t*, 33*f*
defects on parts and, 22, 22*t*
Surface mount technology (SMT), 459
Surface plot, three-dimensional, 380, 381*f*
Surface scratches, example of discrete random variables, 41
Symmetric continuous distribution, Wilcoxon signed-rank test and, 239

T

t confidence interval, 181, 234
t distribution, 180, 180*f*
percentage points of a, 180, 180*f*
t probability density function, 180
t-random variable, 329
t-statistic
curvature sum of squares and, 406
to test that a coefficient is zero, 394
t-test
allows testing against one- and two-sided alternative hypotheses, 292
calculated *P*-value for a, 216*f*, 217
comparison to the Wilcoxon rank-sum test, 264
lower-tailed, 217
normality and the, 219–220
one sample, 216, 220
paired, 264–265
pooled, 254
type II error probability for the, 220
upper-tailed, 217
use of, 289–290
T, temperature, 12
Tabular CUSUM, 468–471, 469*t*, 470*f*, 470*t*
Tampering, definition of, 10
Taylor series expansion, first-order, 13
Taylor's expansion, 64
Temperature (*T*), 12
condensate, 5, 6, 8, 8*f*, 9*f*
example of continuous random variable, 41
reboil, 5, 6, 8, 8*f*, 9*f*
Test
ANOVA, 322
chi-square, 223
contingency table, 232–234
difference in two means and a nonparametric, 261–264
for significance of regression, 291, 291*t*, 322–325, 323*t*, 324*t*
general regression significance, 326
goodness-of-fit, 229–232
homogeneity, 234
hypothesis and lower-tailed, 209
hypothesis and two-tailed, 208–209
large-sample, 215, 225–227, 248
Mann-Whitney, 261
one- and two-sided, 212–213
partial *F*-, 327
partial or marginal, 325
power of a statistical, 201
probability and sensitivity of a medical, 40
P-value for a *z*-test, 209*f*
sign, 235–238
Wilcoxon rank-sum test, 261–262
Wilcoxon signed-rank, 239–240
z, 209–210, 209*f*
Test sample, 5
Test statistic
based on standard normal distribution, 208
calculated to determine the *P*-value, 225
computing from sample data, 196
for a paired *t*-test, 264
for ANOVA, 322
for difference in means, variances unknown and not assumed equal, 257
for difference of two population proportions, 274
for slope, 289
for zero correlation, 301
goodness-of-fit, 230
hypothesis tests on variance and using a, 222
normal approximation for sign, 237
part of procedure for hypothesis tests computations on the, 206
reference distribution for a, 247
standard normal distribution is reference distribution for a, 247
Testing, hypothesis, 7
on the mean of a normal distribution, variance unknown, 218
Tests
on equality of variances from two normal distributions, 270
on individual regression coefficients and subsets of coefficients, 325–329
on the difference in means of two normal distributions, variances unknown and equal, 254–255
on the difference in means, variances known, 248
two one-sided (TOST), 241
Tests of hypotheses, for a single sample, 193–243
Tests on the mean
of a normal distribution, variance unknown, 215–222
variance known, 210
Theorem
Bayes', 39–40
central limit, 15, 152–154, 153*f*, 154*t*
De Moivre's central limit, 73–74
Thermal science, 3
Thinking, statistical, 2–5
Tier chart, 452
Time
example of continuous random variable, 41
observation of processes over, 9–12
Time-oriented plot, 9
Time sequence plot, 140–141
type of graphical display, 140
Time sequences, 140
Time series, 140
Time series plot, 141*f*
data versus time in a, 9, 9*f*
provides more information than dot diagram, 9*f*
stability easier to judge in a, 9
Time to failure, exponential distribution used in reliability studies for device, 86
Time-weighted charts, as alternative to Shewhart control chart, 462–471
Tolerance chart, to assess process capability, 452
Tolerance diagram, 453*f*
Tolerance interval
as a type of interval estimate, 172
for a normal distribution, 191–192
purpose of, 172
Tolerance limit, natural, 453
Total corrected sum of squares, 291
Total probability rule
definition of, 35
for multiple events, 36
Total sum of squares, describes total variability of data, 355
Transformation
the Jacobian one-to-one, 121
variance-stabilizing, 297
Treatment
each factor level is a, 353, 354
two different conditions and if either produces an effect is called, 246
Transformed variables, 303
Treatment effect, 355
Treatment mean square, 356
Treatment sum of squares ($SS_{\text{Treatment}}$), definition of, 356
Tree diagram, 21*f*, 23–24
and conditional probabilities, 33, 33*f*
Trend, 140

Trial
Bernoulli, 55, 57
with only two possible outcomes (Bernoulli), 51
Trials, independent of each other in a random experiment, 51
True regression model, 283
True value, of the population, 130
Two-factor factorial experiments, 382–388, 382*t*
Type I error
definition of, 197
probability of, 197, 201
Type II error (β)
and choice of sample size, 211–215, 220–222, 224, 227–229, 249–250, 259–260, 272, 276
definition of, 197
finding the probability of, 211–212, 212*f*
probability for the *t*-test, 220
probability, 198–201, 199*f*, 200*f*
sign test and, 237–238

U

U chart
an SPC problem-solving tool, 459*f*, 471
center line and upper and lower control limits on a, 458
control chart for defects per unit, 458
U.S. Environmental Protection Agency, limits for PM10 and ozone set by the, 96
Unbalanced experiment, definition of, 359
Unbiased estimator, 156–158, 158*f*, 287–288
Unbiasedness, 161
Unbounded number of outcomes, 45
Uncontrollable factors, also called noise factors, 377
Uniform distribution, 92
Union
definition and notation for, 21
probability of a, 30
type of set operation, 21
Union of events, 29
Uniqueness property, 123
Unreplicated, as a single replicate of a 2^k design, 402
Upper control limit (UCL), on a control chart, 437, 443

V

V mask, used to make limits for CUSUM, 467–468, 467*f*
V, viscosity of a material, 12
Value
expected, 47, 49, 72, 110, 356, 384
integer and noninteger, 46
Variability
all unmodeled sources of (ϵ), 12
definition of, 3
gasoline mileage and, 3
how sample variance measures, 128–129, 129*f*
making decisions when there is data, 2
or scatter of data, 4
quality improvement and reduction of, 435
sources of, 3
Variable. *Also see* Variables
Bernoulli random, 305
beta random, 92
binomial random, 52, 80, 123
bivariate normal random, 116
categorical, 339
chi-square random, 183
continuous random, 41, 67, 121
continuous uniform random, 72
dependent, 312
discrete random, 41, 121
dummy, 340
Erlang random, 87, 88
F-random, 346
gamma random, 87
geometric random, 55, 460
hypergeometric random, 60, 62
independent random, 98, 105–106, 110, 113
indicator, 339
lognormal random, 91
negative binomial random, 58–59, 59*f*
normal random, 63, 73, 123
predictor, 281, 284
qualitative response, 305
quantitative, 339
random, 4
regressor, 284, 312
response, 305
selection of, 341
standard normal random, 75, 77*f*, 91
t-random, 329
Weibull random, 89, 90
Variables
candidate, 342, 348
concept of independence and multiple random, 109–110
discrete random, 43
distribution of a random subset of, 109
expected value of a function of two random, 110
nonlinear quadratic relationship between, 144*f*
nonzero correlation of two random, 113
positive/negative/zero correlation between random, 115
potential relationship between, 144
random, 40–41
regressor, 349
transformed, 303
Variables control chart, 439
Variance. *See also* Mean and variance
analysis of, 351–374
analysis of (ANOVA), 382–383
conditional, 104
confidence interval on a, 184
definition and equation for, 47
equation for estimator of, 320
error, 283
estimator of, 287
hypothesis tests on the, 222–223, 222*f*
normal distribution and tests on the, 223
of a linear function, 118
of a point estimator, 157–158
of a Poisson random variable, 65
one-sided confidence bounds on the, 184
pooled estimator of, 253–254
sample, 151
tests on the standard deviation of a normal distribution, 222–224
Variance components, ANOVA and, 365–368
Variance inflation factor (VIF), 349
Variance known
confidence interval on the mean of a normal distribution, 172
confidence interval on the mean, 173
one-sided confidence bounds on the mean, 176
sample size for specified error on the mean, 175
Variance model, components of a, 365
Variance of the error term ϵ, estimating, 319–321
Variance table, analysis of, 291, 291*t*
Variances
computer calculations for equal, 272
equal, 272
known, 245–261
test on equality of, 270
unknown, 253–259
unknown and equal, 254–255, 260
unknown and not assumed equal, 257, 261
Variation
cause of disruption in voice communication system, 19, 19*f*
need to understand, quantify, and model the type of, 18
Variations, process is affected by types of, 11
Venn diagram, 22, 23*f*, 35
for several mutually exclusive events, 30, 31*f*
represents mutually exclusive events, 23*f*
represents sample space with events in it, 22, 23*f*
Viscosity (*V*), of a material, 12
Voltage, example of continuous random variable, 41

W

Warning limits, 443
Weibull distribution, 89–90
  mean and variance of the, 90
  used to model failure time of a device, 86
Weibull probability density function, 90*f*
Weibull random variable, 89, 90
Weight, example of continuous random variable, 41
Western Electric zone rules, 443, 443*f*
Whisker, on a box-and-whisker plot, 139, 139*f*
Wilcoxon, Frank, 239
Wilcoxon rank-sum test, 261–262
  comparison to the *t*-test, 264
  description of the, 262–263
  sometimes called the Mann-Whitney test, 261
Wilcoxon rank-sum test statistic, and normal approximation, 263
Wilcoxon signed-rank test
  and symmetric continuous distribution, 239
  preferable to the sign test, 240
  similar to *t*-test for symmetric distributions, 240
  ties in the, 240
Wilcoxon signed-rank statistic, normal approximation and, 240
Wilcoxon signed-rank test, 239–240

X

X chart, individuals control chart, 450
$\overline{X}$ control chart, 440, 440*f*, 442*f*
  computer construction of, 449
  control chart for averages, 444–446
X matrix, often called the model matrix, 317
χ^2 distribution, 182–183, 183*f*

Z

z-test, 209–210, 209*f*
  on a proportion, software package sample size calculations for a one-sample, 228–229
Zero correlation, test statistic for, 301

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.